# Fitness-based mating: A systematic analysis of a new preference model

Von der Fakultät für Mathematik und Informatik
der Universität Leipzig
angenommene

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

DOCTOR RERUM NATURALIUM
(Dr.rer.nat.)

im Fachgebiet

Informatik

vorgelegt

von Diplom-Mathematikerin Susanne Schindler

geboren am 22.02.1979 in Halle a. d. Saale.

Die Annahme der Dissertation haben empfohlen:

1. Prof. Dr. Wolfgang Banzhaf (Memorial University of Newfoundland)
2. Prof. Dr. Joachim Hermisson (Universität Wien)
3. Prof. Dr. Jürgen Jost (Max Planck Institut für Mathematik in den Naturwissenschaften)

Die Verleihung des akademischen Grades erfolgt auf Beschluss des Rates der Fakultät für Mathematik und Informatik vom 24.01.2011 mit dem Gesamtprädikat CUM LAUDE.

**Wissenschaftlicher Werdegang**

**1997–2002**
Studium der Wirtschaftsmathematik an der Hochschule für Technik, Wirtschaft und Kultur (HTWK) in Leipzig und der Oxford Brookes University in Oxford, Großbritannien
*Abschluss:* Diplom der Mathematik (FH)

**2002–2004**
Studium der Mathematik an der Universität in Leipzig
*Abschluss:* Diplom der Mathematik

**2005–2010**
Promotionsstudium an der Universität in Leipzig unter der Betreuung von Prof. Dr. Jost am Max-Planck-Institut für Mathematik in den Naturwissenschaften

**Bibliographische Daten**

Fitness-based mating: A systematic analysis of a new preference model /
Susanne Schindler. - 2010. - 135 S. : Ill., graph. Darst.
Leipzig, Univ., Diss., 2010

**Acknowledgements**

I wish to thank Prof. Dr. Jürgen Jost for the enriching collaboration and his support.

Furthermore, I would like to thank Prof. R. Bürger for valuable comments on an earlier version and Dr. A. Grüning, Dr. N. Bertschinger, Dr. K. Ghazi-Zahedi, P.-Y. Bourguignon, Dr. N. Ay, Dr. E. Olbrich, H. Wilhelm, Dr. R. Stollhoff, E. Noeth, F. Bauer, and S. Horn for their constructive proofreading.

Also, I am indepted to the library of the *Max-Planck-Institute for Mathematics in the Sciences* with its comprehensive and manifold assortment of scientific literature.

Leipzig, den 27. Juni 2010
Susanne Schindler

# Contents

# 1 Introduction

## 1.1 Contribution to research

In this thesis, I will propose a new mate preference that is based on ecological performance and name it *fitness-based mating*. Individuals that express this mate preference choose primarily fit partners. The fitness of an individual here corresponds to that individual's expected contribution to the number of (viable) offspring produced in a single mating. Some of the details will be addressed later, but the present definition should suffice for the purpose of this introduction.

The mate preference will be modelled for haploid, diallelic populations. Individuals will be distributed across two niches, and genomes will be simplified to two loci. The first locus will be subject to natural selection, and the 2nd-locus genotype will give the strength of the mate preference.

I will examine under which conditions the preference for the fittest spreads in a population and compare fitness-based mating to assortative mating—which is a preference for like mates—on grounds of their biological plausibility and their evolutionary success. To the best of my knowledge, no such examination of fitness-based mating exists.

In the following, I will outline why mate preferences are significant, which role fitness-based mating takes in the context of evolutionary theory, and how the model of fitness-based mating relates to prior work in the field. Finally, I will give a theoretical and biological motivation of fitness-based mating and I will discuss its possible impact on sympatric speciation processes.

**Role of mate preferences.** Natural populations exhibit a non-random mating behavior which can impact the genotypic and phenotypic clustering in a population. It is assumed that mate preferences causing non-random mating behavior play a role in speciation, and more precisely in sympatric speciation. Under *sympatric speciation*, subpopulations diverge geno- and phenotypically even though they populate the same habitat and could mate with members of other subgroups. Sympatric speciation is distinct from allopatric speciation where external causes inhibit gene flow between subpopulations. Mate preferences are intrinsic, i.e., population-internal mechanisms that can reduce or inhibit gene flow between incipient species. Mate preferences can thus help to understand sympatric speciation processes.

**Natural selection and sexual selection.** Fitness-based mating is at the borderline between natural and sexual selection, because it conditions a mating decision on the ecological performance of the mate. The distinction between natural and sexual selection is motivated by the two components of reproductive success: adaptation to environmental conditions and adaptation to mating conditions.

Observations show that natural selection and sexual selection can interact negatively. For instance, the horn production in some beetles increases their reproductive success, but reduces at the same time the size of organs in the vicinity of these horns [21]. The cost of ornaments or brightness of plumage of some birds can increase not only predation risk [74], but also cause costs in time and energy for maintenance [97]. Also in mammals, the degree of the sexual dimorphisms in males is positively correlated with a higher mortality rate [73].

Natural and sexual selection can also interact positively as is the case in species where physical strength and health is correlated with mating chances (e.g., in elephant seals [25] or sticklebacks [61]). As a direct consequence of my model setup, sexual selection and ecological selection will work in the same direction under fitness-based mating, because both favor fit individuals.

In the fitness-based mating model, the probability to express the mate preference is coded in individual genes and is subject to indirect selection. The strength of sexual selection has therefore its source in individual mate preferences.

**Prior work.** Most studies investigate natural and sexual selection separately. The ecological adaptation has been thoroughly analyzed and even applied to practical problems. Engineers have implemented artificial evolutionary operators in optimization or search algorithms [40, 32]. Also the study of sexual selection alone has received much attention. The most prominent biological examples for sexual selection are exaggerated male ornaments or weapons as are seen in peacock tails or deer antlers. These phenomena can be explained by runaway selection [28] as well as by costly signaling[1].

Meanwhile, also various studies dealing with the interplay between ecological and sexual selection have been published, as for instance by Lande [51]. A prominent model incorporating both selection schemes is that of Dieckmann and Doebeli [16] which led to an enormous amount of follow-up studies. Here, disruptive ecological selection emerges due to competition between population members and induces a selection pressure for assortative mating. Sexual selection enters by means of assortative mating.

The model I will present is a population-genetic model that allows analytic treatment. It focuses on the effect of fitness-based mate choice. The first gene does not exclusively affect ecological performance, but has also a

[1] Costly signaling is a theory which developed from the handicap principle [103] and economical signaling theory [86].

reproductive significance, because it is the object of the mate preference. In this approach, ecological and sexual selection work hand in hand, as ecological selection triggers sexual selection, which in turn then leads to better ecological adaptation. As said before, there is no work examining the role of fitness-based mate choice in detail.

**Theoretical assumptions of the fitness-based mate choice.** Why should individuals discriminate mating partners? Mate preferences are evolutionary useful when they increase the reproductive success of individuals, i.e., a higher expected number of offspring. I will assume that a trait quantifies how much a mating with the trait-carrier increases the reproductive success of the mate. The trait in question is further assumed to be genetically determined. It may depend on the environment which genotype leads to a better trait value. It will be a decisive assumption that the trait value in question can be reliably identified so that individuals can correctly affirm which mating partners possess the "right gene" in a specific environment. Examples are the size and quality of a nuptial gift, containing precious nutrition for the mate or the offspring, or of a breeding territory a male can secure. It could also depend on the environment which trait value is better. For instance, in one environment, it may be better to be big and strong, and in another, small and swift. However, I then need to assume that individuals know which is the better trait value in the given environment when they choose a mate. I also need to assume that the observed trait value is a reliable indicator for the value of the underlying gene. That is, I will not touch the issue to which extent individuals can simply pretend to possess good genes without actually having them. Clearly, in a situation where a trait value decides about mating success, there will be evolutionary pressure to cheat and fake good genotypes,[2] but I shall assume that individuals can reliably assess the value of their potential mating partners. The nuptial gift or the breeding territory seem to represent good examples.

**Biological justification for fitness-based mating.** Fitness-based mating is applicable to all sexually reproducing animals that can actively choose among several mates. So plants for example, which have to rely on transport of pollen by wind, insects, or other creatures are excluded. The ability of choosing among several mates is essential for applying fitness-based mate choice. But in nature, prevalent mating decisions are not made by means of mate preferences, but by mechanisms of behavioral, ethological, or mechanical sort. Yet, looking closer at mating rituals often reveals elab-

[2] The display and the survey of physical strength has led to the theory of honest advertisement and the theory of costly signalling or the handicap principle (see review [44]). Both state that weapons, ornaments, bright plumage etc. indicate the physical constitution. The handicap principle in addition suggests that the production and the maintenance of indicating traits is costly [103]. An example is the study of passerine birds, which examines the costs for plumage brightness in terms of mortality rates [74].

orate tests that apparently aim at choosing more healthy, strong, caring, or technically skilled partners.

Criteria of choosiness could have been developed through direct or indirect advantages which certain mates convey [47]. *Indirect advantages* are those in terms of viable and fertile offspring, which in turn have high reproductive success. They are "indirect" because offspring have high expected reproductive success. The chosen mate might pass on certain properties to the offspring, which increase not only its survival, but also its mating chances. These properties can be the display or the use of some ornament or weapon, or winning a fight against competitors or outcompeting rivals in other ways. If one mate increases the reproductive success of the other mate, it is said that the first mate provides *direct advantages* for the second mate. Examples for direct advantages are qualitative territory with rich resources, paternal care, protection abilities, or food supply. Fitness-based mating is a preference for partners which provide direct and not indirect advantages. This is because indirect advantages may not by itself confer any superiority in raising offspring. But if that is established as a mating criterion in a population, then winning individuals should be expected to produce winning offspring, which then leads to higher numbers of offspring in subsequent generations. However, choosing the fittest clearly provides primarily direct advantages.

How do individuals evaluate the fitness of potential mates? Most species perform species-specific mating rituals in which males compete for mating opportunities, be it through fights, competitions, visual presentation, dances, or harem keeping. Thus, the fitness value could be a genetically determined performance in a mating application round (for instance, wrapping and carrying of mating gifts, harem leadership, or dung ball rolling; see further examples in section 2.4.2 or [2] p. 463).

Fitness-based mating is not fixed to one specific trait value, it is an unspecified preference for the fittest. This is a critical assumption because it means that absolute fitness values are visible by all individuals. Also those elements that immigrated from other niches are assumed to readily detect the fittest genotype even though in their home niches it might be a different genotype. Since different ecological conditions lead to different performances in mating contests, the preferred trait also depends on the geographical position and its ecological peculiarities. For example, depending on open or closed land, different strategies for harem defense prove different efficiency. Dances, display, or competition results vary with temperatures, day times, background noises or background patterns, etc. Thus, also mating preferences can depend on ecological conditions.

**Speciation.** A selection-migration equilibrium is the balance between the opposing forces of selection and migration, where selection is originally understood as natural selection. Selection favors different adaptations in different niches, while migration counteracts the fixation of the best adapted genotypes by constantly mixing the gene pool. The selection-

migration equilibrium is quantified by the equilibrium frequency of the ecologically best adapted genotype. The more migration, the lower the selection-migration equilibrium is. The stronger selection is, the higher the selection-migration equilibrium is. Introducing a mate preference for ecologically well adapted partners increases the selection pressure for fit individuals. Fitness-based mating amplifies natural selection and thus leads to a higher selection-migration equilibrium in the niches. In this way, a polymorphism—that is, the co-existence of several genotypes in a population—is amplified by maintaining a higher equilibrium frequency of well adapted genotypes in the respective niches.

A polymorphism is a preliminary stage of speciation, because subpopulations correspond to groups that share similar phenotypes and probably also similar genes. Speciation is completed when those subpopulations are reproductively isolated. As an indicator of incipient speciation, polymorphic populations are monitored, for instance, in Asian elephants [27], and receive attention in theoretical biology. As I will demonstrate, fitness-based mating leads to a stable polymorphism and thus provides the basis for the development of reproductive isolation. Even more, the polymorphism under fitness-based mating is stronger than under assortative or random mating. In this way, fitness-based mating could provide the basis of further divergence of subpopulations and/or emergence of prezygotic[3] isolation mechanisms, i.e., reinforcement.

## 1.2 Outline

The thesis is structured in the following way: I will introduce biological terminology and population-genetic background in the first two sections of chapter 2. The third section includes general remarks on modelling speciation. The fourth section discusses and compares assortative mating and fitness-based mating on grounds of their biological occurrence and practical realization.

Chapter 3 contains an overview of the literature and highlights where the fitness-based mating model extends existing approaches for the understanding of mate choice and its implications. In chapter 4, I will introduce fitness-based mating step by step. I will formalize it and I will analyze the population dynamics in mathematical terms. The analytical predictions are supplemented by computer simulations. Implementation details and outcomes are presented in chapter 5. The thesis closes with a summary in chapter 6.

[3] prezygotic = prior to fertilization

# 2 Background

## 2.1 Definitions and terminology

This chapter presents relevant biological background. Not all biological concepts have so far been clearly defined or agreed upon. So, I will provide the definitions that are used in this thesis. This introduction is centered around the kingdoms of animals and plants and concentrates on sexually reproducing organisms. I will use material from [90], [9], and [11]. First, biological details and genetics are introduced; concepts of evolution, selection, and adaptation follow. Special attention is given to the fitness and species concepts as well as to the definitions of a polymorphism, a niche, the reproductive isolation, and speciation.

In the second section, relevant population-genetic results are briefly introduced. The third section provides some general remarks about the modeling of speciation and the fourth section presents a comparison of the two mate preferences: assortative and fitness-based mating.

### 2.1.1 Genetics

*Inheritance* is the transmission of characteristics from living creatures to their offspring, as long as the information for the expression of these characteristics is physical. One carrier of transmission is the gene. A *gene* is a physical unit of inheritance. The terms *genetic material* or *genome* denote the whole set of inheritable information of a cell.[1] They become manifest in the DNA; strands of DNA form the *chromosomes*. The exact physical position of a gene on the chromosome is called *locus* (plural form: loci).

In multicellular organisms, cells are classified into *somatic cells* and *germ cells*. There are two differences. Most somatic cells[2] contain the whole genome and germ cells contain only one half of the genome[3]. This structural difference is associated with the different functions that those cells have. The transmission of genetic material to the subsequent generation proceeds via germ cells. Somatic cells, in contrast, do not pass genetic information to the offspring, but only to other cells of the same organism via cell division.

*Sexual reproduction* is the offspring formation caused by the fusion of two germ cells. Since two parents are involved in offspring production,

[1] I will not touch the issue of epigenetic inheritance, because it is not relevant for fitness-based mating.

[2] Exceptions are red blood cells which have no nucleus.

[3] It is not just any half of the genome, but a specific half. This will become clear when the formation of germ cells is outlined later in this section.

each parent can contribute half of its genome.[4]

If somatic cells have more than one set of chromosomes, the organism is called *polyploid*. If a cell has exactly one set of chromosomes, it is called *haploid*; if it has two sets, *diploid*. Germ cells of diploid organisms are haploid.

The *expression* of genetic material is a multistep procedure and results in the formation of proteins. An expressed gene influences the measurable characteristics of the organism. The *phenotype* is the sum of physiological, morphological, or behavioral characteristics of an organism that can be observed. Sometimes the phenotype denotes just one part of the observable characteristics like a specific trait. It is crucial to understand the difference between genotype and phenotype. The *genotype* is the exact genetic material of an organism and is partially passed to the offspring. The phenotype results from the gene expression and is determined by the interaction of the genotype and the environment. The phenotype itself is not passed to offspring.[5]

Genes are placeholders and can take various values. Such altering gene forms are called *alleles*. One allele is a specific nucleotide sequence that a gene can have. If a gene is expressed, different alleles can result in different phenotypes, but several alleles can also result in the same protein. Usually, if there are two allele classes for one gene, these allele classes are labeled by big and small letters, $A$ and $a$ for example. One allele class captures all alleles that give rise to the same protein.

In diploid organisms, each chromosome appears twice in a cell and specific alleles on the same locus may differ. If the very same allele is present on both genes at the same locus, the gene is called *homozygote*, otherwise *heterozygote*. So, genotypes $AA$ and $aa$ are called homozygote and $Aa$ heterozygote. We assume that maternal and paternal chromosomes are not discerned. Hence, heterozygotes $Aa$ and $aA$ are not distinguished. When speaking of zygosity, it is always done so with respect to one specific gene.

### 2.1.2 Inheritance procedure

Passing genes on to offspring is a procedure that includes the replication of genetic material (meiosis), the combination and shuffling of genes (recombination), and the alteration of genes (mutation).

A *mutation* is an alteration of genetic material usually induced by copy errors during DNA-replication. There are so called point mutations, where one nucleotide is inaccurately replaced. Frameshift mutations refer to copy mistakes when a nucleotide is inserted or deleted. If a mutation has no effect on the phenotype, the mutation is called silent. Mutations can also alter the ordering of the genes. Gene sequence mutations occur due to translocation, deletion, duplication, or inversion of several genes. Somatic mutations are not passed to the offspring. Germ cell mutations,

[4] There can be slight asymmetries, for instance, mitochondria are transmitted through mothers only, but these asymmetries are details not needed for this thesis.

[5] Exceptions are mechanisms of epigenetic inheritance.

in contrast, are passed on to subsequent generations. Relevant for this thesis is only the inheritable case and thus the term mutation is restricted to alterations of genetic material in germ cells. The *mutation rate* gives the proportion of nucleotides of a gene that change within a certain time. The mutation rate can vary considerably between different genes.

A further alteration of genetic material is performed by recombination. *Recombination* is a crucial step of the production of germ cells (meiosis) and leads to a new arrangement of the genetic material. It differs from mutation in that it does not delete or insert any genetic material.[6] Recombination exchanges genetic material between two DNA-strands of two chromosomes. It takes place via crossing over during meiosis.

The *meiosis* is a specific cell division process which halves chromosome number and recombines maternal and paternal genetic material. Firstly, maternal and paternal chromosomes are replicated and thus, double in numbers. Then maternal and paternal chromosomes pair off and their genetic material recombines via crossing over, where the chromosomes exchange chunks. This procedure is organized by enzymes. The sites of fracture that determine begin and end of the chunks to be exchanged can occur at any location along the chromosomes. As a consequence, more distant genes are more often recombined than spatially closer genes. The phenomenon that some genes do not assort independently is called *linkage*. Linkage can be caused by spatial proximity on the chromosome, but also by selection pressures. In population genetics *linkage disequilibrium* is a measure for the deviation from random allele distribution. The recombination is followed by two cell divisions splitting the doubled and recombined genetic material into four gametes. *Gametes* are germ cells and have only one chromosome set, they are haploid. Gametes are essentially equivalent to germ cells and this thesis does not distinguish between them. During sexual reproduction, two gametes fuse with each other into a diploid cell. The cell created by gamete fusion is called *zygote*.

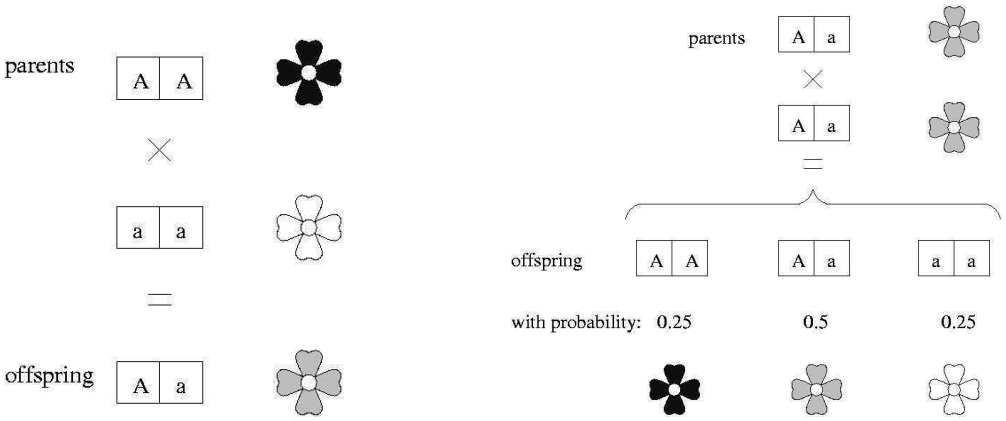
Mendel discovered certain inheritance rules by observation and breeding experiments only [60]. At his time the molecular basis of inheritance was still unknown, but his discoveries revealed its discrete character. The *Mendelian inheritance scheme* is a set of statistical rules that heredity of diallelic genes in diploid organisms usually obeys.[7] There are three Mendelian rules: The uniformity rule, the segregation rule, and the rule of independent assortment:

- The *rule of uniformity* predicts that different but homozygotic parents have solely heterozygotic offspring (see figure 2.1a).

- The *segregation rule* predicts that heterozygote parents get hetero- and homozygotic offspring, each with 50% probability. Which ho-

[6] In molecular biology, also the introduction of foreign DNA into a genome, for instance insertion of virus DNA by enzymes, is referred to as recombination.

[7] There exist phenomena like jumping genes or autosomal killers that cause inheritance to deviate from Mendelian rules.




(a) Rule of Uniformity

(b) Rule of segregation

Figure 2.1: The first two Mendelian rules

mozygote results, is again random with 50% probability each (see figure 2.1b).

- The third rule, the *rule of independent assortment*, treats two traits and is applicable only when the genes coding those traits are situated on different chromosomes. The location on different chromosomes ensures that both genes are inherited independently and therefore, the genes are unlinked. The rule predicts that the probability that the offspring has a specific genotype combination is the product of the probabilities for each trait according to the first two Mendelian rules.

### 2.1.3 The concepts of evolution, selection, and adaptation

I will work out the differences between evolution, selection, and adaptation in this section. The three concepts are often used in synonymy and this section is motivated by providing clear and distinct definitions which are subsequently used in this thesis.

**Evolution.** The probably most rudimentary definition for evolution is the following: *Evolution* is the change in time of allele frequencies in the gene pool of a population. This definition is free of any semantic connotation and so, it excludes any philosophical arguments about improvements, optimization, or directedness. The notion of a population is essential for evolution, because evolution in the sense of the preceding definition can only manifest itself within a population. A *population* is a group of creatures of similar kind that share the same geographical area and are biologically and physically able to reproduce with each other. The *gene*

*pool* of a population is the set of all genes with their current alleles of all individuals of this population. Speaking of a gene pool always refers to a specific population. *Evolutionary factors* are all processes that lead to a change of allele frequencies in a gene pool.

Evolutionary thoughts can be dated back to Greek philosophers such as Aristotle. Later, fossil records were used to either support existence of evolution or to reject it. Nowadays, they are interpreted in favor of the hypothesis of evolution, but have been used in previous centuries to support other ideas like catastrophism[8] and natural theology[9] [9]. During the last decades, results from molecular biology [67] and developmental genetics [31] have confirmed the evolutionary view on life.

**Selection.** Evolutionary theory is based on the concept of selection, although allele frequencies can change due to various mechanisms, for instance simply due to inheritance processes. A definition for selection as elementary as that for evolution is the following: *Selection* is the existence of differences of reproductive success in dependence on the phenotype. The *reproductive success* of an individual is the number of its offspring that survive to reproductive age. Selection is seen as the result of all internal and external circumstances that determine survival and reproductive success. The number of offspring capable of reproduction is hard to determine in practice, therefore, approximations are used like the number of eggs, the number of matings or successful matings, or the number of progenies. It is useful only if compared to the reproductive success of other individuals. Restricting the comparison to members of one population is necessary within the framework adopted here. Therefore, also the concept of selection pertains to a specified population.

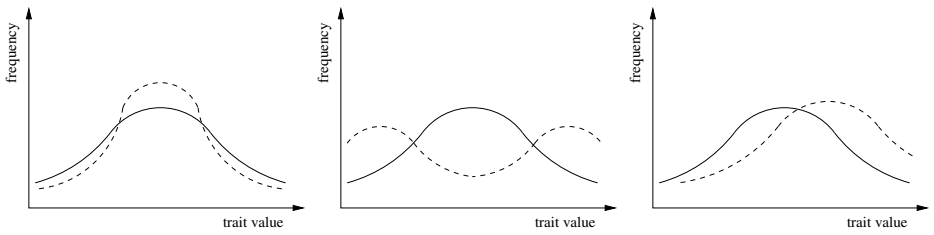
Selection induces changes in phenotype distribution. The phenotype can be measurable on a numerical scale like height, weight, tail length, etc. Other traits like limb construction, teeth arrangement, blossom architecture constitute challenges for a well thought-out numerical scale which allows not only the plotting of distribution, but also meaningful comparison between phenotypes. Selection can change the mean, variance or higher moments of the phenotype distribution.

Preconditions for selection are threefold. Firstly, a population has to carry some variation in a specific trait. Secondly, there has to be a connection between the trait and the reproductive success. Thirdly, the connection between traits of parents and offspring has to be at least partially independent of the environment.

It is very tempting to personalize selection with phrases like “selection acts”, “selection favors”, or likewise “a selection for”. All three phrases

[8] Catastrophism is also named cataclysm theory. It assumes that repeated catastrophe events erased all life forms in certain areas with following resettlement of species that survived outside of the catastrophe area. Other representatives of catastrophism proposed the erasure of all life forms on the world with subsequent creation acts.

[9] Natural theology embraces beliefs that species are everlasting and separately created. Natural theologists aimed to reveal divine plans by mere observation.




(a) Stabilizing selection (b) Disruptive selection (c) Directed selection

Figure 2.2: Classification of selection schemes regarding their effects on the phenotype distribution. Solid lines mark the phenotype distribution of an earlier generation, dashed lines mark the distribution in one of the subsequent generations.

are a synonym for existing differences in reproductive success plus a subliminal presumption that a specific property equips its carrier with relatively higher reproductive success. Also this thesis identifies selection as a state (=differences in reproductive success) with the circumstances causing these differences.

Selection can be classified into various overlapping categories, for instance into stabilizing, disruptive, and directed selection. This categorization is done with respect to the changes to the phenotype distribution caused by selection. If the extreme phenotypes have lower reproductive success such that they are eliminated from the population, one speaks of *stabilizing selection*, see figure 2.2a. Here selection lowers the variance of the phenotype distribution. The larkspur plant in the Rocky Mountains serves as an example. It mostly blossoms brightly, but a tiny fraction of albinos blossom pale or white. Those albinos are constantly produced by mutation and are as fecund as non-albinos. Still, the ratio of albinos reproducing to all albinos is significantly lower than those of non-albinos. This is caused by pollinator preferences. Hummingbird and bumblebee pollinators constantly discriminate against albinos and prefer brightly blossoming plants [98].

One speaks of *disruptive selection* when the two extreme phenotypes have higher reproductive success than the intermediate phenotypes. In this case, the variance of the phenotype distribution is increased, see figure 2.2b. Apparently, disruptive selection acts on several mating strategies for males [35]. A salmon population serves as an example. Early maturing small males compete with late maturing big males for access to females. Both groups of males use different insemination strategies. Dependening on the body size they either sneak or fight with competitors [34]. Small males tend to sneak a copulation, whereas big males, which need longer for maturing, tend to fight with competing males. Other examples are the persisting polymorphism[10] of different beak sizes in African finches [84],

[10]Polymorphism will be defined in section 2.1.6. For the present purpose it suffices to see

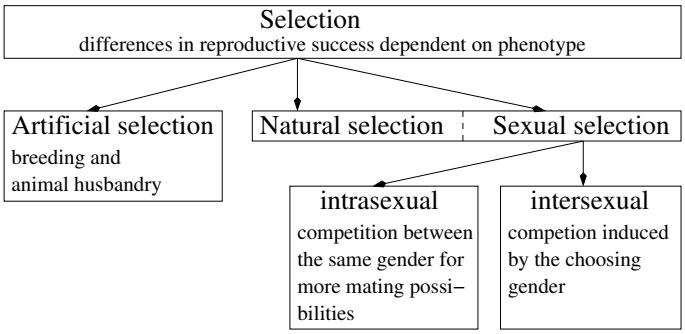



Figure 2.3: Selection schemes distinguished according to their causes.

territorial vs. non-territorial ruffs [95], and dominant vs. non-dominant orang-utans [93]. The under-representation of intermediate traits, in the case of beak sizes, or the under-representation of mixed strategies, in the case of behavior, points to disruptive selection.

*Directed selection* refers to situations when the mean of the phenotype distribution is changed in a specific direction over subsequent generations, see figure 2.2c. Examples are the beak size evolution in finches [33], the shape of oral jaws in cichlid fishes [1], or see [37] for more examples.

Selection schemes can also be classified into frequency-dependent and -independent selection. If the reproductive success of a trait depends on the trait distribution in the population, one speaks of *frequency dependence*. If the reproductive success is independent of it, the selection is called *frequency independent*. Frequency-dependent selection is separated into positive and negative frequency dependence. In the first case the most frequent phenotype is favored and in the second the rarest phenotype is favored. The warning signals of noxious organisms are an example for positive frequency-dependent selection. They advertize by acoustic, olfactory, or visual signals their poisonous effect on predators. Since the most common warning signal is most likely to be learned by predators, individuals displaying the most common signal have higher survival chances than others with different signals. The flu virus in humans or other pathogens is exposed to negative frequency-dependent selection: Viruses of a common strain experience an immune response in most individuals, whereas viruses of a rare strain are hardly exposed to an immune response.

Another classification of selection types is distinguishing the supposed causes of the differences in reproductive success. Figure 2.3 gives a graphical representation. The first kind of selection is *artificial selection* like animal husbandry or breeding, where humans control the reproductive success. *Natural selection* captures all ecological mechanisms which determine the survival and influence the reproductive success in terms of ecological adaptation. *Sexual selection* is a synonym for the differences in the

polymorphism as a stable coexistence of several different morphological traits in a population.

reproductive success due to differences in mating success. A distinction is drawn between the increase of mating success due to the outcompetition of same-gender population members and the increase of mating success due to attractiveness on the choosing gender. The first type of selection is called *intrasexual selection* and captures, for instance, male rivalry or prehensile organs for the prevention of females' escape or assault of other males during copulation. The second type of selection, the differences in mating success caused by differences in attractiveness on the choosing gender, is called *intersexual selection* and captures all mate preferences. The boundary between sexual selection and natural selection is blurred, since a trait can increase both survival chances and mating success.

Another selection classification concerns population elements. Also more abstract entities, not only biological populations, can constitute a *unit of selection.*[11] If, for instance, organisms, demes, species, or alleles are grouped to a population, also organisms, demes, etc., can be units of selection. Depending on that level, one speaks of *individual selection, group selection, interdemic selection*, *species selection*, etc. Of course, the definition of reproductive success has then to be adapted in a natural way, because the number of offspring is different for an individual, a group, or an allele. In the case of alleles, for instance, reproductive success is defined as the number of daughter cells or offspring the allele has been passed on to. It is equivalent to the genetic contribution to the next generation.

Common belief is that competition causes selection. Since all creatures are capable of multiplying in numbers [56],[12] a population quickly experiences natural borders. Darwin's theory of natural selection states that the phenotype that uses environmental and individual conditions most efficiently has the highest reproductive success [13]. This leads to the following paragraph:

**Adaptation.** The next important concept treated is adaptation. *Adaptation* is the process which is triggered by the interaction between the population and the environment and leads to the intensification of some trait. One might call it directed trait development. It is important that these traits must be inheritable and the intensification has to increase survival and reproduction chances, thus reproductive success. In this way, adaptation is closely related to directed selection. Differences are marginal, but while directed selection neutrally describes the situation where more intense traits have higher reproductive success, the notion of adaptation has additionally a subliminal connotation explaining the reasons for these differences.

What is trait intensification? Again it is futile to apply this notion without a population and without monitoring it over several generations. Intuitively, intensification of a trait like spine erection, increasing pathogen

[11] The existence of several units of selection is controversially discussed, see for instance [85].

[12] Even though Malthus [56] deduced his principle for the human species, it is generally transferred to all species, which is supported by empirical evidence.

resistance, tail growth, etc. can easily be captured. In mathematical terms, it corresponds to a reduction of the variance in the phenotype distribution and a shift in mean over time. The intensified trait is called an *adaptive trait* and the intensification an *evolutionary response* to a specific environmental condition. The evolutionary response was caused by the selection of preexisting traits together with the effects of other genetic operators such as mutation and recombination. Each life form can be seen as a set of adaptive traits which enable survival and reproduction in its normal habitat. But not all existing traits of the life form need to be adaptive or have been adaptive in the past.

All three concepts, evolution, selection, and adaptation, are connected to changes of genotype and phenotype distribution of a population. While evolution is the general description of some change in allele or genotype distribution, selection is one particular cause of that change. The adaptation is the specific resulting phenotype distribution.

All three concepts refer to populations which are monitored over a time span of several generations. Not individuals evolve, but populations. Evolution is only measurable when a specific trait or an allele in a population is followed over some generations. The heritability of that trait is an essential aspect. Likewise, not one individual is selected, but the traits are selected that increase the reproductive success of its carrier in comparison to other individuals. Adaptations are the result of change in phenotype distribution over multiple generations due to selection.

### 2.1.4 The concepts of fitness and reproductive success

One of the most difficult and controversial concepts of evolutionary biology in this thesis is the one of *fitness*. The notions of fitness and selection are closely related. While selection is the existence of differences in reproductive success that are ascribable to traits, fitness describes how frequencies of traits are expected to change [85]. A trait that is fitter than a second trait implies a higher reproductive success of the first trait. This implication persists irrespective of the exact physical explanation causing the difference in reproductive success. In this way, fitness is an explanatory placeholder for the physical causes that determine the reproductive success of a trait and its carrier [85].

Fitness is supposed to be amenable to quantification. In the literature, fitness can either refer to a propensity to produce a certain number of viable and fertile offspring, and it can then be quantified, for instance, as the expected number of such offspring under certain environmental conditions and population states, or it can measure the actual number of such offspring produced, possibly over many generations, see for instance [28], pp. 142–143 or [46].

When I will speak of the *fitness* of an individual in a given environment (niche), I refer to the quantitative value of some trait that is correlated with the expected number of offspring arising from a single mating. In

particular, the use of the term "fitness" is non-reflexive in the sense that it does not include the mating strategy. I then need to distinguish between fitness and reproductive success. Reproductive success—which is the overall number of offspring of one individual—is measurable after reproduction, whereas fitness is correlated to an inherent property of each individual. For instance, fitness could quantify ecological performance such as resource allocation or usage, or breeding and rearing abilities. Fitness influences reproductive success, but this correlation is not necessarily positive or strong. For instance, ecological performance and reproductive success can be weakly or negatively correlated as the following two examples will show. Elongated peacock's tails increase mating success [71], but are rather obstructive in daily life or on escape from predation. Here the mating success, which is in this case a reaction to female preferences, is negatively correlated to ecological performance. For humans in industrialized countries, there is a low correlation between eyesight and reproductive success.

Fitness yields the expected offspring number resulting from a single mating. Reproductive success yields the expected overall offspring number. So, reproductive success is composed of the survival in a given environment and the mating success through attractiveness on potential mates. It is not always clearly distinguishable whether an element's reproductive success is due to its own abilities using environmental conditions efficiently or due to its attractiveness. However, both causes depend on each other, since the most ingenious property is not passed to the offspring if its carrier fails to mate, and likewise, the most attracting property is useless if its carrier or its offspring do not survive due to lacking efficiencies.

### 2.1.5 The concepts of species, niche, and reproductive isolation

**Species.** A *species* is a taxonomic category in the classification of animals and plants. For most eukaryotes[13] it is the taxon that is defined most precisely. Still, there are various definitions in use, each more adequate in one context and less in a different context.

The most commonly used species concept is the *biological species concept* introduced by Mayr [59]. According to that concept, a *species* is a reproductively isolated population that shares a common gene pool and a common niche. I will adopt the biological species concept in this thesis, even though the biological species concept is applicable to recent life forms only, it does not account for viable and fertile hybrids which often occur in plants, and the reproductive isolation is hard to test in practice. The three conditions that a population has to meet to be considered as a species will be discussed in the sequel. The first condition is a common gene pool. A *common gene pool* implies some homogeneity of the genetic material over all population members. It is somewhat subjective to define threshold values for genetic variances, especially because some species dis-

[13] Eukaryotes or eukaryonts are cells with nucleus and membranes.

play an enormous variance in genetic composition and others have a gene pool with very low variation. In general, a common gene pool corresponds not only to a relatively homogenous genetic material across population members, but also to a commonly shared lineage up to some ancestors in the past. Molecular biology and stochastics help in clarifying this issue [76].

**Niche.** The second condition for being a species is the sharing of a common niche. A *niche* could be described as the entirety of physical and biotic conditions as well as interactions with other populations. A niche is also described as a functional position with its peculiarities and needs in an ecosystem.

Various debates arose about the niche definition. This is due to the two existing different predominant approaches. The species-centered view and the environment-centered view mark extreme points on an abstract scale for niche definition and the discussion circles around where to place the niche definition. The species-related definition of a niche takes interdependency of species and ecosystem into account. In that framework a niche cannot be *taken*, but *emerges* by the interactions between species members and their environment, such that one can identify species with the niche. The environment-centered niche framework defines a niche as the sum of all biotic and abiotic life conditions. Those can be used by several different populations. Usually, one argues with the exclusion principle that a niche is finally *taken* by organisms that use life conditions most efficiently.

As already mentioned, the species-centered view and the environment-centered view for niches represent opposite ends of a scale. Most other niche definitions rely on a combination of both views and are oriented to varying extent to one view or the other. In cases where populations have such high impact on the environment, such that they create or improve their life conditions, one speaks of niche construction. A prominent example are symbiotic relationships, which can be even obligate to at least one symbiont. Therefore, some ingredients of the species-centered view are necessary for the niche definition. On the other hand, observations show a certain amount of exchangeability of populations in a given geographic position like it is seen in resettlements after extinction events. Hence, also the environment-centered view is justified. In other words: The population might create up to a certain extent more favorable life conditions than were present beforehand, but those conditions cannot be created arbitrarily.

The species-centered definition has—for the use within this thesis—the considerable handicap that it uses the species definition for the niche definition. This leads to a circular definition in which a niche is equivalent to a species and species members occupy a common niche. For that reason, I will use the environmental niche definition.

**Reproductive isolation.** The third condition for a species is the reproductive isolation to other populations. Two populations are *reproductively isolated* if there is no gene exchange between them because of geographical, behavioral, or genetic barriers. Reproductive isolation is classified into *prezygotic* isolation, which includes all isolation mechanisms operating prior to zygote creation, and *postzygotic* isolation, which includes all mechanisms that operate after zygote formation. Among prezygotic isolation rank ecological or spatial isolation when meetings of different population members are inhibited by different habitat preferences, ethological isolation when matings fail due to population-specific behaviour, mechanical isolation when copulation organs are structurally incompatible, and gamete mortality such that fertilization is impossible even with successful insemination. Postzygotic isolation mechanisms include zygote mortality such that despite successful fertilization development terminates, hybrid inferiority, and hybrid sterility such that gene flow into the generation of grandchildren is inhibited. Prevailing isolation mechanisms are of prezygotic kind, which might be due to selection for the avoidance of possibly inferior hybrids to save the costs of lethal or infertile hybrids.

### 2.1.6 The definitions of polymorphism and speciation

**Polymorphism.** A polymorphism originally implies the existence of several different morphological characteristics in a population, but depending on the field that definition varies. For instance in population genetics, a *polymorphism* is defined not only in the phenotype space, but also in the genotype space. In genetics, a polymorphism is the stable coexistence of different genotypes in a population. In biology, it is additionally required that those genotypes give rise to different phenotypes. But in general, a polymorphism is exhibited by population members of the same species. Biological examples for polymorphisms are the following: Garter snake populations consist of subpopulations with different skin patterns. Some snakes are longitudinally striped, which impedes inference of their speed while moving, and some snakes are speckled and blur more easily with the underground [7]. Subpopulations show different antipredator behavior, striped snakes tend to flee, speckled tend to remain motionless ([9], p. 463). Another polymorphism, which is so far only genetically detectable, is reported for Asian elephants. They are classified into two matrilineages, which represent genetic clustering without morphological differences [27].

**Speciation.** Speciation is the process by which new species arise. According to the biological species concept, *speciation* is equivalent to the emergence of reproductive isolation between different subpopulations of the group that was formerly one species. This emergence of reproductive isolation is usually accompanied by the accomplishment of genetic differences (separated gene pools) and phenotype differences (different niches). It is a problem during speciation when to tell incipient species apart, i.e., when are geno- or phenotypic differences sufficiently large. A possibly

| | $x_i, x_j$ | | |
|---|---|---|---|
| $x$ | $A, A$ | $A, a$ | $a, a$ |
| $A$ | 1 | 0.5 | 0 |
| $a$ | 0 | 0.5 | 1 |

Table 2.1: Offspring distribution $P(x|x_i, x_j)$ for haploid individuals. Parents $x_i, x_j$ produce offspring $x$ with the probability $P(x|x_i, x_j)$.

artificial threshold has to be defined to separate the formerly common lineage into two branches.

Three modes of speciation can be distinguished. They are allopatric, parapatric, and sympatric speciation. Originally, they refer to the geographical situation during the speciation process, which influences gene flow between subpopulations in an essential but not exclusive way. Gene flow is a measure for the extent of sexual contact between members of subpopulations. Gene flow can not only be restricted by geographical, but also by temporal or preferential circumstances. Hence, more precisely stated definitions of speciation modes are based on presence or absence of gene flow. *Allopatric speciation* captures speciation processes in which gene flow between incipient species is disabled. If gene flow is possible and not prevented by external circumstances, speciation is called *sympatric*. *Parapatric speciation* refers to somewhat suppressed gene flow between incipient species and therefore, it takes an intermediate position between allopatric and sympatric speciation.

## 2.2 Population genetics

This section presents population-genetic details which are relevant for the fitness-based mating model. At first, the Mendelian inheritance scheme will be formalized, secondly, the Hardy-Weinberg equilibrium will be presented, and thirdly, the linkage disequilibrium will be defined. More comprehensive introductions to population genetics are found in [24] or [75].

### 2.2.1 The Mendelian inheritance

Mendel [60] revealed the discrete character of inheritance which is summarized in the Mendelian rules, see section 2.1.2. As we will need the Mendelian rules for the modeling, I will introduce you to to the formalism that is applied in the sequel. Let $P(x|x_i, x_j)$ denote the probability that haploid parents $x_i$ and $x_j$ produce the haploid offspring $x$. The haploid offspring distribution is set as in table 2.1 and the diploid offspring

distribution is set to:[14]

$$P(xy|x_ix_l, y_jy_k) = P(x|x_i, x_l)P(y|y_j, y_k) + P(y|x_i, x_l)P(x|y_j, y_k) \qquad (2.1)$$

where parents' genotypes are denoted by $x_ix_l$ and $y_jy_k$ and offspring's genotype by $xy$. The first term of equation (2.1) represents the probability that the offspring receives the allele $x$ from one parent, say its mother, and the allele $y$ from its father. The second term gives the probability that the offspring receives $x$ from its father and $y$ from its mother.

In this formalism, $P(xy|x_ix_l, y_jy_k)$ obeys the first and second Mendelian law. Let us check the law of uniformity: Parents have genotype $AA$ and $aa$. Then

$$P(Aa|AA, aa) = P(A|A, A)P(a|a, a) + P(a|A, A)P(A|a, a) = 1 \qquad (2.2)$$

whereas $P(AA|AA, aa) = P(aa|AA, aa) = 0$. The rule of segregation is reflected in $P(AA|Aa, Aa) = P(aa|Aa, Aa) = 1/4$ and $P(Aa|Aa, Aa) = 1/2$.

In a similar way, the formalism is generalized to two loci. Let the second locus take alleles $M$ and $m$. Possible genotypes are then $MM$, $Mm$, and $mm$. I will identify a two-locus genotype by its numbers of $A$ and $M$-alleles. So, $il$ with $i = 2$ and $l = 1$ corresponds to genotype $AAMm$. Offspring probabilities are then denoted by $P_{\text{off}}(mn|il, jk)$. If the two loci are unlinked, for instance when they are situated on different chromosomes, then the probability that an offspring of parents $il$ and $jk$ has the genotype $mn$ can be calculated by

$$P_{\text{off}}(mn|il, jk) = P(m|i, j)P(n|l, k), \qquad (2.3)$$

where $P(m|i, j)$ and $P(n|l, k)$ are defined by equation (2.1). The probability $P_{\text{off}}(mn|il, jk)$ reflects the third Mendelian law, that is the rule of independent assortment, which holds only when the two loci are unlinked.

### 2.2.2 The Hardy-Weinberg equilibrium frequencies

The Hardy-Weinberg frequencies refer to a genotype equilibrium of a randomly mating population (e.g., [36]). Further assumptions are that the gene in question is diploid and diallelic, and that selection pressure and mutation are absent. That means that alleles do not affect fitness and no new allele appears.

Let $p$ denote the relative allele frequency of $A$, then the Hardy-Weinberg frequencies of genotypes $AA$, $Aa$, and $aa$ are $p^2$, $2p(1-p)$, and $(1-p)^2$, respectively. The Hardy-Weinberg frequencies are attained already after one generation and depend on the allele frequencies only, i.e., not on the genotype frequencies. They are thus equilibrium frequencies, because there is neither selection nor mutation which could change the allele frequencies.

[14] A similar notation has been used in [64]. There, $R_{i,jk}$ and $S_{i,jk}$ denote the probabilities of the production of gamete $i$ by male and female parents with gametes $jk$. Likewise, $R_{ij;kl,mn}$ gives the respective probability when parents and offspring are diploid.

### 2.2.3 Linkage disequilibrium

The distribution of alleles over two loci is not necessarily independent and can thus not always be calculated by the product of one-locus distributions. Let the first gene take the alleles $A$ and $a$ and let the second gene take the alleles $M$ and $m$. The relative gamete frequencies are denoted by $\gamma$. For instance, $\gamma_{Am}$ gives the frequency of gametes with allele $A$ from the first locus and allele $m$ from the second locus. The gametic frequencies can be calculated from the genotypic frequencies denoted by $P$:

$$\begin{aligned}
\gamma_{AM} &= P(AAMM) + \frac{1}{2}\Big(P(AAMm) + P(AaMM)\Big) + \frac{1}{4}P(AaMm) \\
\gamma_{Am} &= P(AAmm) + \frac{1}{2}\Big(P(AAMm) + P(Aamm)\Big) + \frac{1}{4}P(AaMm) \\
\gamma_{aM} &= P(aaMM) + \frac{1}{2}\Big(P(AaMM) + P(aaMm)\Big) + \frac{1}{4}P(AaMm) \\
\gamma_{am} &= P(aamm) + \frac{1}{2}\Big(P(Aamm) + P(aaMm)\Big) + \frac{1}{4}P(AaMm).
\end{aligned} \tag{2.4}$$

The linkage disequilibrium is denoted by $D$ and is calculated by

$$D = \gamma_{AM}\gamma_{am} - \gamma_{Am}\gamma_{aM}. \tag{2.5}$$

If genotype $j$ of the first locus and $k$ of the second locus assort independently, meaning $P(j,k) = P(j)P(k)$, the gamete frequencies are

$$\gamma_{AM} = P(A)P(M),\ \gamma_{Am} = P(A)P(m),\ \gamma_{aM} = P(a)P(M),\ \gamma_{am} = P(a)P(m) \tag{2.6}$$

and thus

$$D \overset{(2.5)}{=} \gamma_{AM}\gamma_{am} - \gamma_{Am}\gamma_{aM} = 0. \tag{2.7}$$

The gametic disequilibrium $D$ quantifies how much alleles differ in their distributions among gametes. If loci assort independently, there is no covariance and therefore, there is no correlation between the loci, thus $D = 0$. The inference in the other direction is not possible: $D = 0$ follows from independence, but independence does not follow from $D = 0$. If the loci do not assort independently, $D \neq 0$.

## 2.3 Modeling speciation

In this section, I will make some general remarks about speciation models. Speciation can be seen as a clustering in an abstract space [54]. Let the genotype space be the set of all possible genomes. Genetic operators such as recombination and mutation can be described as functions on the genotype space. The mutation is then a mapping of the genotype space onto itself, and the recombination assigns two parental genomes

to one offspring genome. The recombination thus maps the product of the genotype space onto itself. For recombination, a mating function is necessary which groups a population into a set of pairs. A pair symbolizes a mating between its elements. Selection acts at the phenotype level—in contrast to mutation and recombination which act on the genotype level—because selection is the result of all ecological and individual circumstances impacting the survival and the reproduction. Thus, a map from genotype space to phenotype space is needed which assigns a phenotype to each genotype. This map needs not to be bijective.

The evolution is then the result of the repeated action of recombination, mutation, and selection. It can be visualized as moving clouds through the genotype-phenotype space. Species that obey the biological species concept ([59], see section 2.1.5) are separate clusters in this space which neither overlap nor merge in the course of evolution. This is because a cluster in genotype space refers to a common gene pool. And a common niche is a cluster in the phenotype space combined with some information about geographical and climate conditions as well as about ecological interactions with other clusters. Reproductive isolation is met when there is no mating between the members of distinct clusters.

A polymorphism can be a step in the process of speciation. A polymorphism is the long-term maintenance of distinct alleles in the population, see section 2.1.6. It is therefore a transitional state for the splitting of one population into two subpopulations with different gene pools. Several models—which are going to be reviewed in chapter 3—try to give conditions for the emergence and the maintenance of a polymorphism. The contribution of this thesis is a new model which leads to a stable polymorphism. I will introduce this model in chapter 4. The difference between polymorphic subpopulations and distinct species is the presence or absence of matings across the clusters they represent in the genotype-phenotype space. Polymorphic subpopulations without reproductive isolation between them do not constitute species. But if they are reproductively isolated, then the polymorphic population has split into species.

Explaining speciation is equivalent to describing a mechanism that splits one cluster into two—or more—clusters without matings across these clusters. Speciation is the result of local interactions (matings) together with ecological conditions, genetical encoding, and inheritance rules. The modeling of speciation in this framework has two degrees of freedom[15]. The first is the definition of the ecological conditions and the second is the definition of a local interaction rule. Inheritance rules and genetic setup are modelled in a standard way, that means by the use of Mendelian inheritance rules, see section 2.2.1, and the use of finitely many loci with several alleles.

An interesting concept is that of Breidbach and Jost [6]. They suggest that speciation and the co-existence of species can be seen as the result

[15] As opposed to robotics where a degree of freedom equals a parameter, a degree of freedom in this context means the possible beginnings of modeling.

of and the balance between diverging and converging trends in the gene pools of populations. The mutation and the selection are captured as diverging forces, because mutation induces noise in the gene pool and selection induces pressures in certain directions of the population's genetic composition. The recombination, in contrast, is seen as a homogenizing force [6] causing the mixture of genetic material and acting as an accelerator for the spread of advantageous mutations. However, recombination on the short time scale is a constant source of variation, because it keeps recombining genetic material. But the steady mixing through recombination leads eventually, on the long time scale, to a homogenous gene pool, provided that mutation is not altering genetic material.

## 2.4 Comparison of assortative mating and fitness-based mating

In this section, I will compare two mate preferences, assortative mating and fitness-based mating, on grounds of their biological justification. Let me first define the term mate preference. It is crucial to distinguish between a mating that results from external constraints and a mating that results from a decision in favor of one mate while rejecting other mates. External constrains are, for instance, a heterogeneous population distribution in space and time or the physical compatibility of the reproduction apparatuses. But a *mate preference* is an internally driven choice for a mate that has certain traits or properties when also other mates without these properties are available. When I will speak of a mate preference then I assume that the preference is genetically determined. The *assortative mate preference* or *assortative mating* then is a choice for similar mates, when also dissimilar mates are available. It is based on both the own and the candidate's phenotype. The *fitness-based mate preference* or *fitness-based mating* is a choice for fit partners when also less fit partners are around. It is based on the candidate's fitness only.

Observations of natural populations show that pairings are not random, but pairs exhibit correlations between trait values. If pairings were random, no correlations should turn up.[16] It is a challenge to find out about the reasons for these correlations. A well thought-out experimental design can help to reveal whether these correlations show up because of a side effect of some mating relevant condition or because individuals exhibit in fact a mate preference. In the following, I will present biological examples where the traits of mating partners show non-random pairings and I will point to possible causes for these phenomena. I will group the examples and their possible causes according to whether mates exhibit more similarities than would be expected under random mating (assortative mating), and to whether fit individuals receive more mating chances than less fit individuals (fitness-based mating).

[16] Correlations can also be caused by limited sample size. In that case they vanish when the controlled set is made larger.

### 2.4.1 Assortative mating

The phenomenon of assortative mating receives attention, because it is thought to play a role in speciation processes. A population can be divided into groups of individuals having a resemblance in a specific trait. If individuals pair assortatively, matings should take place predominantly within those groups. Hence, gene flow between groups of dissimilar trait values is low or absent. The reduction of gene flow or the incipient reproductive isolation induced by assortative pairings (*pairings* as opposed to *matings*) can be a cause or a result of speciation. It is causing speciation if it reduces gene flow between subpopulations to the extent that subpopulations can gather genetical and phenotypical alterations. In its most rigorous form, assortative pairing can induce reproductive isolation. On the other hand, assortative pairing follows from speciation, because the incipient reproductive isolation is a form of assortative pairing.

Assortative pairings are observed in various species, but the reasons of such correlations between mates remain sometimes unknown. I will present some biological examples of assortative pairings in natural populations and their possible causes in the next subsections.

#### Biological examples of matings between like mates

**Strawberry poison frog.** The strawberry poison frog (*Dendrobates pumilio*) is native in Central America. These frogs have eye-catching skin coloration, which has probably developed as a warning for predators to indicate the toxic alkaloids in the frog's skin. The strawberry poison frog is organized in several geographically separated populations which show different skin coloration. A population of the same skin color is called a morph. Two morphs (green and orange skin color) were monitored with respect to their mating behavior [91]. Females were presented males of both morphs and it was recorded how long and how close the females oriented towards the presented morph. The experiments were carried out under two light conditions, once under white light and once under blue light, which makes distinction between orange and green color difficult. Females displayed a significant preference for males of the same morph under white light, whereas they displayed no preference under blue light. This species can thus serve as an example for assortative mate choice.

**Darwin finches.** A population of medium ground finches (*Geospiza fortis*) shows a bimodal distribution in terms of a beak size measure [42]. The beak size measure includes beak length, depth, and width. The population is separated into two groups with small and large beak morphs and it has relatively few individuals with intermediate beak sizes, which might be a response to food availability and competition. Observational data displayed strong evidence for assortative paring [42], which could result from the feeding place being used for recruiting potential mates.

**Flowering times.** Flowering times enforce assortative pairings, because early flowering plants mate with other early flowering plants with a higher probability than with late flowering plants. The same holds for late flowering plants. In this case, the assortative pairing is caused solely by the availability of potential mates and not by mate choice.

**Body size.** Generally, assortative pairings with respect to body size is observed across many animal species. For instance, leaf beetles pair assortatively according to elytron length [8], which is the length of the hardened forewing of insects.

**Humans.** Assortative pairing among humans is reported for similar mental capabilities, physical attractiveness, similar moral concepts, and physical traits as weight or height (see [96] for a review). In most civilisations, age constitutes the decision factor with highest assortment coefficient (e.g., [72]).

**Inbreeding.** Inbreeding is an extreme case of assortative pairing. The difference between assortative pairing and inbreeding lies in the number of similar gene loci. In the case of inbreeding, almost all genes have similar values, whereas assortative pairing is concerned with the similarity of only some traits which are probably caused by only some genetic loci.

**Disassortative pairings.** The opposite of assortative pairings are disassortative pairings. These are pairings in which the partners tend to have distinct phenotypical trait values. A prominent example for disassortative pairing results from a seemingly preferential mate choice for partners with dissimilar major histocompatibility complex (MHC), a gene region influencing the immune system [99].

A study of [41] hints at disassortative pairing in the white-throated sparrow (*Zonotrichia albicollis*) based on dissimilar plumage.

#### Possible causes for assortative pairings

Assortative pairing can result from facilitated mating, as it is the case in matters of body size or of weight, as well as from mate choices which are not externally enforced. I will sketch in the following some causes of assortative pairing.

**Heterogeneity by spatial, temporal, or behavioral availability.** Barriers to gene flow can be induced by temporal, spatial, or behavioral obstacles. If a population is distributed over an area and some traits are distributed heterogeneously, the availability of mating partners can induce assortative pairings.

An example is the so called *ring species* which is a set of geographically distant populations in which incipient populations can interbreed. Due to

the spatial distance, far-flung populations do not interbreed. The populations form a chain of neighboring and interbreeding, and thus genetically related, populations. Populations at both ends of the chain can share the same habitat such that the chain forms a ring. Possibly, the end populations are genetically too distant and do not interbreed. An example is the herring gull species complex, whose reproductive isolation has been investigated in [55] on a molecular basis.

Some species like fruit flies mate on feeding grounds. Thus, the nutrition habits in conjunction with the predisposition of mating at the feeding places induce assortative pairing, such that animals with similar feeding habits tend to mate with each other. The assortative pairing found in the medium ground finches with respect to beak size [42] might be caused by such a mechanism.

Examples for temporal constraints enforcing assortative pairings are flowering times of plants and activity patterns. The impact of activity patterns is most apparent in nocturnal or diurnal creatures, since they virtually never meet even when living in the same area.

**Co-evolved relationships.** Co-evolved relationships between species form a special case for temporal or spatial constraints inducing assortative pairings. Some species are tightly connected to other species as in host-parasite or plant-pollinator relationships. Such co-evolved relationships can condition the mating process of one species on the life cycle of the other species. For instance, the treehopper (*Enchenopa*) lives on different host plants. These plants differ in their life histories[17] and thereby influence the mating time and egg hatching of the treehopper [102]. Treehoppers pair assortatively depending on which host-species they live on. The possible consequences for speciation are investigated in [18]. An example for assortative pairing caused by pollinator preferences is found in [45].

**Prezygotic or postzygotic enforced assortative pairing.** The compatibility of reproduction apparatuses can coincide with assortative pairing or even with reproductive isolation mechanisms. For instance, the assortative pairing in the leaf beetle with respect to size is due to the *"greater ease with which size-matched pairs can achieve intromission"* [8].

In addition to mechanically enforced similarity, also sufficient genetical similarity is required to assure the compatibility on a genetic basis. The embryonic development terminates when the fused germ cells are incompatible.

[17] life history = strategy for course of life, for instance, the age at sexual maturity, the age at death, or the age at other events in life time; investments into own growth or offspring upbringing, number and size of offspring, gender ratio of offspring, or gender change in hermaphrodites, etc.

**Assortative mating.** Assortative mating causes obviously assortative pairings. Since all external causes are excluded, assortative mating requires an awareness of the own phenotype. That awareness is presumably not present in all species. A solution to the problem of missing awareness could be a genetically encoded fixed mate preference. This implies that each group of phenotypically distinct individuals has different genes causing the mate preference. Translated into a model, this would result in a two-allele model in Felsenstein's categorization [26], see section 3.2.3 on page 43 for a definition of two-alleles models.

### 2.4.2 Fitness-based mating

Among all possible mate preferences, assortative mating receives most attention. But according to Darwin's theory of selection, offspring of fit mating partners should proliferate because they are are either numerous or have a high quality[18]. Therefore, it is puzzling why offspring of similar partners should proliferate. Choosing a fit partner should be a more straightforward and a evolutionary successful strategy.

The assumption behind assortative mating is an awareness for similar mates. The assumption behind fitness-based mating is an awareness for fit mates. It is not obvious—and it is beyond the scope of this thesis—to answer what is more difficult: To evaluate the fitness of mating candidates or to evaluate the similarity to mating candidates. That fitness can be evaluated demonstrates, for example, Darwin's reference to a pigeon breeder, who told that in *"pigeons, which pair for life, the female [...] will desert her mate if he is injured or grows weak"* ([14], p. 249). One way how a trait can predict the reproductive success of its carrier is when there is a correlation between the phenotype and its attractiveness for the choosing gender. The more pronounced this phenotype is, the more attractive is its bearer, the more mating chances the bearer has, and the more offspring it will produce. The assumption that the magnitude of the phenotype is positively correlated with the reproductive success of its carrier is thus self-fulfilling. However, a high number of overall offspring produced by an individual does not necessarily imply a high number of offspring from a single mating. But the latter one is that what the choosing mate is interested in. So, a correlation between trait intensity and attractiveness does not serve as a justification for fitness-based mate choice. The mentioned correlation serves only for the intensification of the trait over the course of time.

An other way how a trait can predict reproductive success—and this is the one that is relevant for fitness-based mating—is when the trait in question encodes a performance in a mating contest. Then the attractiveness follows from a better performance in the contest. In this case, the mating contest should test reproduction relevant features, e.g., fertility, fecundity,

[18] In the sense that they are viable and fertile descendants which give birth to further high-quality children.

food supply and parental care for the future brood, hunting or defensive abilities, territory quality or decreased pathogen transmission (see [47], p. 339 where direct advantages for choosing females are listed). In these cases, a good performance implies a high expected offspring contribution in a single mating thus, it implies a high high fitness value, see section 2.1.4 for the definition of fitness. Contests preceding mating, as well as that individuals having a good performance in these contests receive more mating offers, are often observed in nature and are the rationale behind fitness-based mating; I will give some examples and possible explanations below.

#### Biological examples of matings where fit mates are preferred

I will present some mating rituals in biological species which hint to a selection of fit mates. But besides the selection of physically and behaviorally well adapted mates, a mating ritual can also realize the reproductive isolation to any other species that employs a different mating behavior. Thus, mating rituals could explain fitness-based mating as well as assortative mating at the same time.[19] I will give some examples of mating rituals that are interpreted as either a competition for mating partners or the choice for very fit mating partners.

**Mating dances.** Flies of *Drosophila melanogaster* exhibit an extensive mating dance prior to mating [87]. It involves the circling and the pursuit of the female by the male. If the male keeps up for a certain time, the female stands still and cooperates in mating ([9], p. 434). This might constitute a selection for physically able and healthy males, because males have to keep up for certain duration.

**Male rivalry.** Male rivalry is present in several species and is interpreted as a competition for females. There is a number of mechanisms for outcompeting other males in primates, either before, during, or after mating [48]. Those mechanisms take shape, for instance, in physical dominance or sperm competition. In apes, the set of criteria for a physically and mentally healthy male might be its sensibility for receiving signals from receptive females, may they be acoustic or olfactory [48].

**Harem keeping.** Gorillas (*Gorilla*) are an example where dominant males have a harem in which they monopolize females. They virtually exclusively inseminate the females of their harem. However, the females have a set of responses to turn down mating attempts. Colorful patches, tail length, symmetry of structure, properties of male calls as well as other morphological features can influence the female mating cooperation [48]. The females can thus circumvent the monopolization by one male, which

[19] Clearly, if a mating ritual is based on a sexual dimorphisms such that members of different sexes look different, then assortative mating is not applicable and serves neither purpose.

can influence the success or failure of a takeover attempt by a new group leader. Dominance or leadership in a group can be taken by physically strong male gorillas which are able to protect and lead the group.

A further example are red deer (*Cervus elaphus*). Males defend their harem by physical fights. Fighting success is positively correlated with reproductive success. Field studies providing observation data [10] were supported by DNA-analysis testifying the observed paternities [69]. Also among sea lions (*Mirounga angustirostris*), physical strength correlates with mating success [52, 25].

**Mating gifts.** In some species of *Empididae* flies, the ritual of mating gifts is practised. Even though most species are vegetarians, males catch some small insect prey, wrap it in silk and fly with it into a mating swarm [49]. This swarm consists of dancing males carrying each such a balloon present. Catching and carrying of such balloons is a sign of physical ability and health. Females fly into those swarms, choose a male, accept the mating gift, and mate. In some species of *Empididae*, females cooperate in mating as long as they suck on the prey of the mating gift. The fruitier or bigger the mating gift is, the more probable is the successful insemination and the more it has required the hunting, wrapping, and transportation skills of the male.

**Territories.** Male hummingbirds (*Eulampis jugularis*) occupy and defend territories with nectar-producing plants, on which hummingbirds feed. Since the mating occurs after chasing the female around which feeds in the male's territory [101], a male capable of defending a territory with a larger amount of nectar-producing plants will attract more females. Hence, physical abilities are indirectly favored.

**Leks.** Other examples in birds for favoring dominant males are all species exploiting lek-based mating rituals. The leks are traditional mating grounds where the males gather together and females choose among them. The more central the position in the circular shaped lek is, the more attractive is the male for females [39], but the more the male has to defend its position. Therefore, attracting a somewhat higher number of females involves physical strength.

**Supplies for offspring.** Male dung beetles (*Onthophagus-binodis thunberg*) produce, prior to mating, a dung ball which must be guarded and defended against other males [12]. They roll the ball and if an attracted female follows, they mate and the female lies the fertilized eggs into the dung ball. Both bugs bury the ball on which their offspring feeds after hatching. The bigger the ball is, the more strength is necessary to roll it, but the longer and the more offspring can feed on it.

**Outer attraction.** The darkness and the length of a male lion's mane (*Panthera leo*) indicate testosterone level and hitherto fighting success. It influences male-male competitions and female choice [100].

The secondary sexual adornments of mandrills (*Mandrillus sphinx*), which is a red skin on the face and a fatted rump, signal a male's rank in a group [83]. The offspring number of a male depends on its rank [17]. However, since higher ranked males socialize with females and they practise mate-guarding more often than lower-rank males, it is unclear whether females base their mate choice on those adornments.

So far, examples where the success in mating rituals depends on the ecological niche have not been discussed. But imagine for dung beetles that in niche 1 moist and sticky dung dominates, which can be most easily and esthetically rolled by carriers of allele $A$, whereas in niche 2 dry and crumble dung dominates to which $a$-carriers are specialized. The ball rolling requires different techniques, which leads to a performance difference of dung beetles in different niches, whereas the mating contest (dung ball rolling) is the same in both niches. Transferred to balloon flies, niches might exhibit different wind conditions which affect the mating dance. In populations where males form harems, the ecological conditions like open or covered surroundings might affect the strategies and possibly the necessary physical conditions for harem maintenance. In this way, the traits leading to good performances vary across the niches, and so do the preferred mates.

#### Possible causes for fitness-based pairings

The presented examples support the existence of the signaling of mating relevant information. The mating rituals can be interpreted as preference for males that display signs of high reproductive success in one single mating trial. The set of features that give rise to the belief of high offspring numbers differs strongly between species. Besides fitness-based mating, there are several other reasons that could cause the heterogeneity in the number of mating chances across different individuals. I will outline the most prominent of them in the following.

**Physical dominance.** The physical strength of one mate, be it applied or threatened with, can hinder further matings of its partner. This exclusion of rivals is practised in harems, hierarchical organized groups, or via mate guarding.

**Passive attraction.** A mating decision can be the result of active choice or passive attention that a sender of mating relevant signals receives [68]. While active mate choice includes a rejection of other mates in favor for the one chosen, passive mate choice happens when only mates exhibiting the most intense stimulus are chosen [68]. For example, a louder mate call

spreads over a larger range. More candidates can thus hear and react to the call, which gives its sender more mating options. The theory of sensory exploitation [77] is a similar approach. Here, sexual selection favors mates that use more effectively biases in the sensory system [77]. For instance, the female preference in a frog species for males with lower-frequent calls coincides with a bias of their basilar papilla tuning towards lower-than-average frequencies [78]. In practice, the distinction of active mate choice from passive attention requires thoughtful experimental design. In the case of passive attraction, the favor of mating relevant traits is not caused by a preference but by cognitive requirements.

**Mate copying.** Mate copying [19] is an economic strategy to choose among candidates in little time. In extreme cases, just one active mate choice is drawn in a population and all others imitate and choose the "approved" mate. In experiments, the mate preferences of female guppies can even be reversed when seeing the formerly rejected male in company with another female [20].

**Fitness-based mating.** Obviously, the choice of fit mates leads to pairings with fit partners. The fitness of an individual is its expected contribution to the number of (viable) offspring produced in a single mating. The point of fitness-based mating is that an individual profits in an evolutionary sense from a mating with another individual that, in the given environment, shows a value of that trait that leads to a higher expected number of offspring than other values of that trait. Of course, the actually produced number of offspring will depend on both mating partners, but for any individual, it is advantageous to find a good mating partner in order to increase its own reproductive success. In that sense, individuals may strive to find fit mating partners. This may be a reciprocal situation for both mating partners involved, and this will constitute one of the issues of the subsequent formal analysis in chapter 4.

# 3 Previous theoretical studies

The idea of a genetically controlled drive for fitness-based mating has to my knowledge not yet been examined before. Nevertheless, I would like to use this chapter to list existing work and highlight their relevance to the fitness-based mating model. It will be mostly an overview over models for sympatric speciation and models treating the interplay of natural and sexual selection. Each of these impacted deeply the field, triggering a wealth of further papers on the topic. This overview is by no means complete, but is rather tailored for comprehending the motivation of this thesis' work. A more comprehensive overview is found in [29].

## 3.1 Overview of the field development

The maintenance of a polymorphic population is a highly interesting and non-trivial question. A polymorphism is formalized in population-genetic terms as the stable co-existence of several alleles in a population. But which forces cause a stable polymorphism? Hardy and Weinberg (e.g., [36]) formalized the inheritance process and showed that in absence of selection the existing genetic variance is maintained. When selection is introduced and favors a homozygotic genotype, then the allele out of which the homozygote consists will go to fixation. When the heterozygote is favored, a co-existence of several alleles can result.

Levene [53] has found conditions where the heterozygote is not favored by selection, but a polymorphism develops nevertheless. So, polymorphism can be caused by inheritance processes as well as by ecological selection. Another possibility are mate preferences, which are frequently found in nature and might relate to speciation, either as a cause or as a result. Gene flow is restricted in this way not extrinsically, but intrinsically. So consequently, mate preferences were introduced (e.g., O'Donald [66]). The majority of these studies treats the preference for similar mates, so called *assortative mating*. The connection to matings among species-members only, hence *similar* mates is obvious. Assortative mating represents a prezygotic isolation mechanism. Many contributions (for instance, Dickinson and Antonovics [15], Moore [62], Udovic [92], Felsenstein [26]) elucidated the impact of assortative mating on polymorphism or even reproductive isolation.

Subsequently, an interest arose in sexual selection which possibly counteracts ecological selection. I will summarize the work of Kirkpatrick [50], Spirito [88], and Sanderson [79] to exemplify the trend and sketch the findings.

The awareness arouse that others than the assortative mate preference could lead to speciation. Those preferences were gathered under the name *preferential mating*. I will present the work of Servedio [81] which compares assortative and preferential mating to test their potential to cause reproductive isolation.

Population-genetic models are analytically tractable and allow for a clear distinction between cause and effect. This clarity is achieved at the cost of over-simplification and neglection of finite-size effects. A new approach called *adaptive dynamics* applies dynamical systems theory to the description of evolutionary processes. It is a new framework in which the combined effect of ecological, sexual, and frequency-dependent selection can be analyzed. The foundation was led by Geritz et al. [30]. A computational model by Dieckmann and Doebeli [16] reproduced the predictions of adaptive dynamics and will be presented.

A crucial question remains: How do mate preferences evolve at all? Instead of assuming that a fixed fraction of the population exhibits them, later models examined the individual benefits from indulging in mate preferences. The work of Matessi et al. [57] serves as an example for these studies.

## 3.2 Polymorphism in a population

### 3.2.1 Conditions for polymorphic equilibria

**Levene, 1953.** A stable polymorphism is maintained in a population over a single niche when heterozygotes have the highest absolute fitness value. Generalizing this situation to more than one niche, heterozygotes should be superior to both homozygotes in any single niche. But Levene [53] demonstrated that multiple niches, between which individuals move freely, give rise to additional configurations which lead to a stable polymorphism. He analyzed a 1-locus, diallelic, diploid population which is distributed over several niches. The sequence of the genetic operators is the following: Individuals settle in random niches where they experience a selection pressure. The probability of survival is proportional to the fitness value. Individuals that survive mate randomly across all niches. Offspring replace parents and settle in random niches. The selection and recombination operators are repeatedly applied.

Let $n$ be the niche index. The population fractions $c_n$ are such that $\sum_n c_n = 1$. The absolute fitness values of genotypes[1] $AA$, $Aa$, and $aa$ in niche $n$ are $w_n$, $z_n$, and $v_n$, respectively. The relative fitness values of the homozygotes $AA$ and $aa$ with respect to the heterozygote $Aa$ are $W_n = w_n/z_n$ and $V_n = v_n/z_n$. The conditions for a stable polymorphism

[1] The original notation for the alleles are $A$ and $A'$. For the sake of easy comparison between the models, I will rename the alleles to $A$ and $a$.

| | $AA$ | $Aa$ | $aa$ |
|---|---|---|---|
| niche 1 | $1+K$ | $1+K$ | 1 |
| niche 2 | 1 | 1 | $1+k$ |

Table 3.1: The fitness values in Maynard Smith's model [58].

are

$$\sum_n c_n/W_n > 1 \quad \text{and} \quad \sum_n c_n/V_n > 1. \tag{3.1}$$

If there is only one niche, these conditions are necessary and sufficient. If there are two niches, the conditions are only sufficient [53].

The multiple-niches model itself corresponds to a model with migration at a high rate, namely $p_m = 0.5$. Reducing the migration rate $p_m$ or introducing a habitat preference, as Maynard Smith [58] has done, might ease the threshold for the polymorphism.

### 3.2.2 Polymorphism through habitat choosiness

**Maynard Smith, 1966.** Levene remarked that his model is not realistic and suggested that "*individuals move preferentially to niches they are better fitted for*" [53] instead of randomly choosing a niche. Maynard Smith [58] extended Levene's model for two niches by introducing such habitat preference.

Maynard Smith's model distinguishes between females and males. Males move freely between the niches, but females migrate from the niche where they are born to the other niche with a probability of $(1-H)/2$. Females remain in the niche of birth with a probability of $(1+H)/2$. The parameter $H \in [0,1]$ is a measure for the habitat preference of all females. Setting $H = 0$ implies no habitat selection and coincides with Levene's model for two niches, whereas $H = 1$ implies a strong habitat bonding of females. The parameter $H$ thus lowers the migration between niches.

The absolute fitness values in Maynard Smith's model are reproduced in table 3.1 and depend on the parameters $K$ and $k$. The current relative frequencies of allele $A$ born in niche 1 and 2 are denoted by $p_1$ and $p_2$, respectively. The population sizes are "*separately regulated in the two niches*" [58] which means that each niche can support a certain number of elements. Selection acts separately on both niches. The equilibrium condition that $p_1$ remains constant over two generations leads to the following equation:

$$\frac{k-K}{Kk} = \frac{4S^2 - 8S + 2 + H(1-2S)^2}{2-H} \qquad \text{equation (3) in [58]}, \tag{3.2}$$

where $S := (p_1 + p_2)/2$.

For the two cases $H = 0$ and $H = 1$, Maynard Smith has given analytical descriptions and conditions for a stable polymorphism. If $H = 0$, which is

| couple | mating probability |
|---|---|
| $AA, AA$ | $\alpha P(AA) + (1-\alpha)^2 P(AA)^2/R$ |
| $AA, Aa$ | $2(1-\alpha)(1-\beta)P(AA)P(Aa)/R$ |
| $AA, aa$ | $2(1-\alpha)(1-\gamma)P(AA)P(aa)/R$ |
| $Aa, Aa$ | $\beta P(Aa) + (1-\beta)^2 P(Aa)^2/R$ |
| $Aa, aa$ | $2(1-\beta)(1-\gamma)P(Aa)P(aa)/R$ |
| $aa, aa$ | $\gamma P(aa) + (1-\gamma)^2 P(aa)^2/R$ |

Table 3.2: The mating probabilities with $R = 1 - \big(\alpha P(AA) + \beta P(Aa) + \gamma P(aa)\big)$ in O'Donald's model [66].

equivalent to no habitat preference, the condition for a stable equilibrium is [58]:

$$-1 < \frac{k-K}{Kk} < 1, \tag{3.3}$$

and the upper bound coincides with Levene's condition $\sum_n c_n/V_n > 1$. If $H = 1$, the condition is [58]:

$$-1.5 < \frac{k-K}{Kk} < 3. \tag{3.4}$$

Maynard Smith introduced a second locus on which strict assortative mating acts. The second locus takes the alleles $B$ and $b$. The allele $B$ is dominant, hence there is no phenotype difference between $BB$ and $Bb$-types. Matings take place among $bb$-homozygotes and among $BB$- and $Bb$-types. By means of a numerical example, Maynard Smith showed that a polymorphism at the second locus evolves as well as a linkage disequilibrium between the two loci [58].

### 3.2.3 Polymorphism through mate preferences

**O'Donald, 1960.** O'Donald [66] examined the frequency evolution of heterozygotes in 1-locus models with assortative mating. He proposed two models. One model treats the case when each of the three genotypes[2] $AA$, $Aa$, and $aa$ is distinguishable[3]. A fixed fraction of these genotypes mates assortatively: $\alpha$ of the $AA$-homozygotes, $\beta$ of the heterozygotes, and $\gamma$ of $aa$-homozygotes. The relative frequencies of genotypes $AA$, $Aa$, and $aa$ are denoted by $P(AA)$, $P(Aa)$, and $P(aa)$. The mating probabilities are reproduced in table 3.2.

[2] O'Donald named the two alleles in his 1-locus model $A$ and $B$. I will change the notation into $A$ and $a$ to emphasize that both alleles belong to the same locus, to distinguish clearly from other 2-locus models where allele $B$ occurs at the second locus, and ease the comparisons of the models presented.

[3] This means that there is no dominant allele suppressing the expression of an other allele in a genotype.

There is no mutation and the allele frequencies remain constant, but the distribution of genotypes changes. O'Donald calculated the relative heterozygote frequency as the root of a quadratic equation and gave numerical examples. In principle, low assortative mating, i.e., low $\alpha$, $\beta$, and $\gamma$, leads to a population where almost half of it is heterozygous, which would be similar in a random mating situation. The bigger $\alpha$, $\beta$, and $\gamma$ are, the lower the fraction of heterozygotes is, because assortatively mating homozygotes produce only homozygotic offspring, and assortatively mating heterozygotes have descendants half of which are homozygotes. The overrepresentation of homozygotes is thus a polymorphism—in terms of allele co-existence—which is not caused by a high number of heterozygotes.

The second model treated by O'Donald is also interesting. Here, allele $A$ is dominant and $\alpha = \beta = \gamma$. Again, there is no mutation and the allele frequencies remain constant. Since $AA$ and $Aa$ are indistinguishable, it is of particular interest at which frequency the heterozygote $Aa$ is maintained in the population under the cover of its homozygotic phenotype. O'Donald calculated in an elegant way the equilibrium frequency of $Aa$, as well as the rate of approach to the equilibrium. Both can be given as functions of $\alpha$ and the initial relative frequency of allele $A$. Again, small values of $\alpha$ lead to approximately 50% of heterozygotes and this fraction is decreasing with higher $\alpha$.

Assortative mating leads to an accumulation of homozygotes. Random mating leads to an accumulation of heterozygotes. O'Donald's credit was the quantification of the contrary effects of random and assortative mating.

**Dickinson and Antonovics, 1973.** Dickinson and Antonovics investigated the influence of assortative mating on the population composition, especially the potential to prevent genetic mixing at an ecological locus that influences the fitness value of the individual but is not relevant for the mating procedure.

Dickinson and Antonovics presented the results of deterministic computer models for two niches with migration and for elements with one, two, and three loci [15]. Dickinson and Antonovics examined several functions for the second locus such as causing dominance, selfing, contributing additively to the fitness value, or being the gene assortative mating acts on, but I will focus on the latter model which leads to reproductive isolation.

Individuals have two loci; the first locus takes alleles $A$ and $a$, the second locus $B$ and $b$. Disruptive selection acts on both loci; it favors alleles $A$ and $B$ in the first niche and alleles $a$ and $b$ in the other niche. Fitness values are reproduced in table 3.3. The fitness value of a combination of the first-locus genotype $g_1$ and second-locus genotype $g_2$ is given by multiplication: $f(g_1, g_2) = f(g_1)f(g_2)$. A fraction of males migrates from the niche of their birth to the other niche. Females are immobile and do not migrate.

Dickinson and Antonovics plotted the relative frequency difference of

| | $AA$ | $Aa$ | $aa$ |
|---|---|---|---|
| niche 1 | 1 | $1-\frac{1}{2}S_x$ | $1-S_x$ |
| niche 2 | $1-S_y$ | $1-\frac{1}{2}S_y$ | 1 |

(a) First locus

| | $BB$ | $Bb$ | $bb$ |
|---|---|---|---|
| niche 1 | 1 | 1 | $1-\Sigma$ |
| niche 2 | $1-\Sigma$ | $1-\Sigma$ | 1 |

(b) Second locus

Table 3.3: The fitness values in the model of Dickinson and Antonovics [15].

| couple | mating probability (not normalized) |
|---|---|
| $AA, AA$ | $P(AA)^2$ |
| $AA, Aa$ | $2\alpha P(AA)P(Aa)$ |
| $AA, aa$ | $2\beta P(AA)P(aa)$ |
| $Aa, Aa$ | $P(Aa)^2$ |
| $Aa, aa$ | $2\alpha P(Aa)P(aa)$ |
| $aa, aa$ | $P(aa)^2$ |

Table 3.4: The mating probabilities in Moore's model [62]. The normalization factor of the mating probabilities is $N_P = P(AA)^2 + P(Aa)^2 + P(aa)^2 + 2\alpha\Big(P(AA)P(Aa) + P(Aa)P(aa)\Big) + 2\beta P(AA)P(aa)$.

the $B$-allele between both niches for fixed selective pressures $S_x = S_y$ and $\Sigma$, fixed proportion $\alpha$ of assortatively mating individuals, and fixed migration rate $p_m$. High selection pressure $\Sigma$, $\alpha$ close to 1, and a low male migration rate $p_m$ led to a pronounced polymorphism at the second locus. The linkage of allele $B$ with $A$ and the linkage of $b$ with $a$ caused an incipient reproductive isolation between both genetic clusters and that fit individuals mate among themselves ($AABB$ in niche 1 and $aabb$ in niche 2) and unfit among unfit (vice versa).

**Moore, 1979.** Moore [62] drew the attention to genotype frequency development when assortative mating is costly, especially for rare genotypes. In contrast to O'Donald's model where all individuals mated, Moore showed that the population dynamics changes qualitatively when finding a mate depends not only upon the frequency of its own genotype but also upon the choosiness.

Moore proposed a 1-locus model for diploid individuals. A couple mates with probability 1, $\alpha$, or $\beta$ depending on whether they have two, one, or zero alleles in common. The relative frequencies of genotypes[4] $AA$, $Aa$, and $aa$ are denoted by $P(AA)$, $P(Aa)$, and $P(aa)$. The not yet normalized mating probabilities are reproduced in table 3.4.

Setting $\alpha$ and $\beta$ to zero, thus having complete assortative mating, demonstrates the population dynamics. The mating probabilities of $AA$, $Aa$,

[4]Moore named the two alleles $A$ and $A'$. For reasons explained in footnote 2 on page 38, I will change the naming into $A$ and $a$.

and $aa$-couples are $P(AA)^2/N_P$, $P(Aa)^2/N_P$, and $P(aa)^2/N_P$. Using the Mendelian rules, see sections 2.1.2 and 2.2.1, matings of the heterozygote distribute 50% of its offspring to genotypes $AA$ and $aa$. Hence the equilibrium distribution will lack heterozygotes and we can shorten the analysis by assuming $P(Aa) = 0$. The mating probabilities for $AA$ and $aa$-couples are then $P(AA)^2/N_P$ and $P(aa)^2/N_P$ and are at the same time the relative offspring frequencies of the next generation. Apart from the case where $P(AA) = P(aa) = 0.5$, the rarer homozygote will become extinct and with it, of course, the rarer allele.

In the remainder of Moore's paper, he investigated the effects of mutation and selection which can counteract the *"'centrifugal' force of assortative mating"* [62]. In the case of mutation, the rarer allele will constantly be reintroduced into the population and Moore raised the question of the frequency equilibrium of both alleles given $\alpha$, $\beta$, and the mutation rate. Moore supplied numerical results, while Gavrilets [29] points to several papers (e.g., [38]) with analytical results. In the case of selection, the question is how strong does selection favor the rarer allele to get a polymorphic equilibrium. The answer is given analytically in [29]: the relative viability fitness of the heterozygote must exceed the sexual selection force $\alpha$ against it.

**Udovic, 1980.** Udovic [92] presented a 2-locus model in a single niche. The first locus takes the alleles $A$ and $a$ and determines the fitness value. Let $j \in \{0, 1, 2\}$ denote the number of $A$-alleles in a genotype. So $j = 0$, $j = 1$, and $j = 2$ correspond to $aa$, $Aa$, and $AA$, respectively. The variable $p$ denotes the relative frequency of the $A$-allele in the population. The fitness values for each $j$ depend on the population composition: $f(j, p)$ denotes the fitness value of genotype $j$ in a population with a proportion $p$ of $A$-alleles.[5] Udovic introduced the functions

$$\begin{array}{rcl} a(p) & = & f(0,p) - f(1,p) \\ b(p) & = & f(2,p) - f(1,p) \end{array} \qquad \text{equations } (2)^* \text{ in [92]} \tag{3.5}$$

which specify the fitness differences between the homozygotes and the heterozygote. The equilibrium frequency of allele $A$ is given by

$$\hat{p} = \frac{b(\hat{p})}{a(\hat{p}) + b(\hat{p})} \qquad \text{equation (5) in [92].} \tag{3.6}$$

The parameter

$$S = a(\hat{p}) + b(\hat{p}) \qquad \text{equation } (9)^* \text{ in [92]} \tag{3.7}$$

is the sum of fitness differences between homozygotes and heterozygotes at equilibrium. The parameter $S$ is a measure of disruptive selection: If both homozygotes are privileged by selection over the heterozygote, then

[5] Udovic denoted the fitness values by $W_j(p)$. I have changed the notation to matching it with that of the fitness-based mating model which will be presented in chapter 4.

$S > 0$. If the heterozygote is favored by selection over both homozygotes, then $S < 0$, and if there are no fitness differences or the values cancel each other out, then $S = 0$.

The parameter

$$M = \hat{p}\frac{\mathrm{d}a}{\mathrm{d}p}(\hat{p}) - (1-\hat{p})\frac{\mathrm{d}b}{\mathrm{d}p}(\hat{p}) \qquad \text{equation } (10)^* \text{ in [92]} \tag{3.8}$$

captures properties of frequency-dependent selection. When fitness values are independent of the genotype frequencies, which is equivalent to the absence of frequency-dependent selection, then $M = 0$. Positive frequency-dependent selection, meaning common genotypes are favored, leads to $M > 0$. Negative frequency-dependence, meaning rare genotypes are favored, leads to $M < 0$.

Udovic analyzed the conditions for a stable polymorphic equilibrium and gave a necessary condition: $S + M < 0$ [92]. The condition $S + M < 0$ expresses that the effects on the stability of a polymorphism caused by disruptive and frequency-dependent selection can compensate each other. For instance, if the heterozygote is less fit than both homozygotes ($S > 0$), the stabilizing frequency dependence of selection has to be sufficiently strong, i.e., $-M > S$, in order to enable polymorphism.

A second locus is introduced by Udovic which takes the alleles $M$ and $m$ and is subject to assortative mating. If an element mates assortatively, it chooses a mating partner which has the same second-locus genotype. The probability to mate assortatively is $\alpha$ for all individuals. Random mating occurs with probability $1 - \alpha$. Udovic gave the conditions for stable polymorphism which is caused by a fixed amount of prezygotic isolation through assortative mating [92] which Gavrilets [29] extended and simplified.

Cases of special interest for speciation are those when homozygotes at both loci are over-represented and genotype combinations of both loci are not arbitrary. Non-arbitrary genotype combinations point to a linkage disequilibrium between the loci. That means, only $AAMM$ and $aamm$ are present in the population, but no $AAmm$ or $aaMM$. Measures for this situation are the heterozygote deficiency $I = 1 - P(Aa)/2p(1-p)$ and the gametic linkage disequilibrium $D$, see section 2.2.3 for a definition of $D$. The measure $I$ quantifies the deviation of the relative heterozygote frequency from the Hardy-Weinberg heterozygote frequency, or in other words, it quantifies the deviation from the expected heterozygote frequency under the assumption that alleles are randomly distributed in the genome. The case $I < 0$ corresponds to an excess of heterozygotes and $I > 0$ to a deficiency.

The analysis including the second locus is analytically cumbersome and Udovic supported it by numerical experiments. He showed the existence and the stability of equilibria of $I$ and $D$ in dependence of $S$ [92]. He detected parameter ranges for these equilibria: $0 < S_b < S_c < -M$. In the first interval $[0, S_b]$, the only stable equilibrium lies at $I = 0$ and $D = 0$. Both conditions imply that the frequency of heterozygotes corresponds

| | $AB$ | $Ab$ | $aB$ | $ab$ |
|---|---|---|---|---|
| niche 1 | $(1+K)^2$ | $1+K$ | $1+K$ | $1$ |
| niche 2 | $1$ | $1+K$ | $1+K$ | $(1+K)^2$ |

Table 3.5: The fitness values in Felsenstein's model [26].

to the product of frequencies of both alleles and that there is no linkage between loci. If $S \in (S_b, S_c]$, then there is additionally one instable equilibrium with $I < 0$ and $D \neq 0$. This corresponds to an excess of heterozygotes. If $S \in (S_c, -M)$, then the stability of the equilibrium $I = 0$ changes from stable to unstable and there is one stable equilibrium with $I > 0$, that is, a deficiency of heterozygotes, and $D \neq 0$. The latter case represents a stable polymorphism at the locus under selection and reproductive isolation between homozygotes, thus, it represents a *"partial premating reproductive isolation"* [92].

**Felsenstein, 1981.** Felsenstein [26] pointed out that speciation can be classified into two types. He called "*these two classes of models of speciation 'one-allele' and 'two-allele' models. The critical distinction between them is whether reproductive isolation is strengthened by substituting the same or different alleles in the two nascent species*" [26]. In other words, consider a diallelic locus that induces reproductive isolation between two subpopulations. If reproductive isolation is strengthened by the fixation of only one allele in both populations, then we have a one-allele model. If reproductive isolation is strengthened by the fixation of one allele in one population and the fixation of the second allele in the second population, then we have a two-allele model.

Furthermore, Felsenstein proposed a two-allele model and he points out that migration and recombination are opposing forces against speciation in two-allele models. He argues that speciation might more likely be caused by a one-allele mechanisms, because speciation benefits in this case from the mixing properties of migration, since one allele is of advantage in both niches.

Felsenstein analyzed a model with three haploid loci. The first locus takes the alleles $A$ and $a$, the second takes the alleles $B$ and $b$, and the third locus takes the alleles $M$ and $m$. The first two loci are subject to disruptive selection. The fitness values of the gametes are reproduced in table 3.5. Assortative mating is acting on the third locus. Each element mates assortatively with a probability of $\alpha$, that means it chooses a mate which has the same 3rd-locus allele $M$ or $m$. Model parameters are the migration rate, the recombination fraction and the selection strength $K$. Felsenstein showed that parameter ranges for speciation are rather small and that migration as well as recombination counteract speciation in two-allele models [26].

**Spirito, 1987.** Spirito [88] examined a partial premating isolation mechanism which is realized by assortative mating. Each genotype $AA$, $Aa$, and $aa$ has its specific mating area, which is a combination of space and time. Mating areas of different genotypes can overlap, controlled by a parameter $z$. The more the more alleles they share, the more their mating areas overlap. In this way, Spirito highlighted a mating procedure resulting in assortative mating without cognitive assumptions on the 2-locus individuals.

A second neutral locus with alleles $B$ and $b$ is introduced.[6] The population starts with individuals equal in frequency of type $AABB$ and $aabb$ only. Spirito examined how fast the linkage between both loci decreases over time depending on $z$ and the recombination rate $R$. In a second model, two populations are linked by migration, one starting with $AABB$, the other exclusively with $aabb$-individuals. The time how fast the linkage decreases is again monitored. The isolation mechanism is efficient only when $z$ is extremely high, i.e., when only very few different genotypes meet (in the one-population model) or when migration rates are low (in the two-populations model).

Gavrilets [29] points out, that Spirito's model [88] is transformed into Moore's model [62] by introducing female preferences.

### 3.2.4 Relevance for fitness-based mating model

Models of polymorphism help to identify conditions under which the genetic variation is maintained when there is a deficit of heterozygotes. The selection of the models presented here shows that fitness conditions, habitat choosiness, and mate preferences can cause polymorphic equilibria and that recombination, migration or mutation, and random mating counteract polymorphic equilibria.

All these models have in common that the mating behavior itself is not subject to selection pressure or evolution. The fitness-based mating model goes in several aspects further as it can build on recent developments in the field: the mate preference is genetically determined and can differ in the expression level. The parameter $\alpha$, that gives the probability for assortative mating of each arbitrary individual in the models listed above except Spirito's [88], is transformed to be the mating probability of a specific individual.

We will see that the choosiness is not costly in the fitness-based mating model. This is an aspect that the fitness-based mating model has in common with Udovic's model [92] and where it differs to Moore's model [62].

[6] The notation chosen by Spirito is $A_1$ and $A_2$ for the 1st-locus alleles and $B_1$ and $B_2$ for the 2nd-locus alleles. See footnote 2 at page 38 for the justification of my renaming.

## 3.3 Opposing forces: Ecological and sexual selection

**Kirkpatrick, 1982.** Kirkpatrick [50] used a diallelic, haploid, 2-locus model to examine the counteracting forces of natural and sexual selection.

The population consists of females and males, where the genotype frequencies are equally expressed in both sexes right after birth. Males then undergo a selection that favors allele $A$ at the second locus[7]. The first locus determines whether the individual, if it is female, mates randomly or prefers male carriers of allele $a$. This allele reduces viability fitness in males. Viability selection acts differently on males and females, because $a$ reduces only male fitness. But Kirkpatrick noted that when $a$ reduces also females fitness, the qualitative model behavior does not change significantly.

Kirkpatrick expressed the genotype evolution in relative allele frequencies and the linkage disequilibrium $D$. Given a specific frequency of the female preference allele, the preferred allele can be maintained in a population even though it conveys fitness disadvantage [50]. This partly supports the theory of runaway selection (firstly proposed in [28]), where the magnitude of a possibly negative trait increases due to its attractivity onto the choosing gender.

**Sanderson, 1989.** Sanderson [79] examined how premating isolation can establish at the presence of postmating isolation.

A 2-locus population is divided into demes which are 1-dimensionally arranged next to each other.[8] Demes are linked by migration from and into neighboring demes. Left-hand situated demes experience selection favoring 1st-locus homozygotes $AA$, right-hand situated demes experience selection in favor of $aa$. Heterozygotic "hybrids" are selected against. The second locus carries only $B$-alleles.[9] This situation gives rise to a *cline*. A cline is a *"geographic gradient in a measurable character, or gradient in gene, genotype, or phenotype frequency"* [22]. Sanderson explored two models. In both, he introduced the 2nd-locus allele $b$ which influences the selection in one model and the mate choice in the other model. The models elucidate non-favorable situations for the evolution of premating isolation mechanisms.

In the first model, $b$-carriers experience a lower fitness disadvantage for heterozygotes $Aa$ but also a slight fitness reduction for homozygotes $AA$ and $aa$. The latter is motivated by the theoretical disadvantage of $b$-alleles outside the zone of hybridization. In the second model, $b$-carriers tend to mate assortatively. Sanderson used computer simulations and found that $b$-alleles in either model rather seldom increase in relative frequency, leading to the conclusion that *reinforcement* has a minor influence in the

[7] The original naming of the 2nd-locus alleles is $T_1$ and $T_2$.

[8] The number of demes is not specified in [79], but in the simulations the number ranges between 40 and 160.

[9] I have renamed 1st-locus alleles $A_1$ and $A_2$ into $A$ and $a$, and the 2nd-locus alleles from $B_1$ and $B_2$ into $B$ and $b$.

speciation process. Reinforcement is the selection of *"characters that reduce matings between races [...] until isolation becomes absolute and new species are produced"* [79].

### 3.3.1 Relevance for fitness-based mating model

Sanderson's model has dealt with genetically determined mate preferences that are under selection pressure. This is important for arguing about the evolutionary advantage of mate preferences and will be an issue for fitness-based mating.

Kirkpatrick's model showed that disadvantages in viability can be compensated by advantages in finding mates. The fitness-based mate preference amplifies genetical differences when natural and sexual selection are working in the same direction, which is in contrast to Kirkpatrick's model where their forces are opposite.

## 3.4 Adaptive dynamics

The adaptive dynamics framework [30] was developed in the 1990s and is an application of the dynamical systems theory to evolutionary processes.

Adaptive dynamics theory follows the distribution of an ecological trait under frequency-dependent selection and aims to give conditions for equilibria. Approximations are based on the following assumptions:

i) the trait is continuous,

ii) the reproduction is clonal, that means, asexual,

iii) mutations are rare, and

iv) mutations have little effect on the continuously varying trait.

The starting point of the examination is a resident population in monomorphic equilibrium and a few mutants. That means, all elements have the same trait value, apart from some rare mutants whose trait differs. Since mutants are rare, it is assumed that their impact on the population which is subject to frequency-dependent selection can be neglected. The growth rate of the mutants can be calculated and it is assumed that this initial growth rate indicates the mutant's frequency development. If the mutant's growth rate is positive, it is assumed that the mutant trait spreads and quickly replaces the resident population, such that the population settles again in a monomorphic equilibrium. Evolution, in the perspective of adaptive dynamics, is seen as a sequence of invasions of mutants. Hence, adaptive dynamics is sometimes referred to as *invasion fitness analysis*. The theory exhibits two time scales, a fast one for the invasion of mutants and a slow time scale for the long-term behaviour as a result of subsequent invasions.

The main interest is the assessment of the stability of the equilibria. The classification into evolutionary stable and unstable equilibrium points are complemented by the categories of convergence stable or unstable traits [23]. A *convergence stable trait* is a trait to which monomorphic populations will always converge and an *evolutionary unstable trait* is one which has a lower growth rate than some mutants.

Especially traits which are convergence stable but evolutionary unstable play a major role in adaptive dynamics. Such traits are called *evolutionary branching points*. Those traits are attracting points of a monomorphic population. Nevertheless, some mutants' growth rates are higher than the growth rate of an evolutionary branching trait. This means, a single mutant will cause the population to move away from the evolutionary branching point. As it also is a convergence stable point, the population will be invaded by mutants whose traits are closer to the branching point. Thus, the population returns back to the branching point in the following evolutionary step. Only the emergence of two mutants at the same time, which differ from the evolutionary branching point $x_0$ in opposite directions say, $x_0+\epsilon$ and $x_0-\epsilon$, can lead to persistent subpopulations which stabilize each other. Since the evolutionary branching point is evolutionary unstable, in each subpopulation, mutants which lie further away from the branching point, but balance each other by traits of the form $x_0 + 2\epsilon$ and $x_0 - 2\epsilon$, have a higher growth rate. Hence, subpopulations move further away from each other in trait space. This is the reason why such convergence stable but evolutionary unstable traits are called branching points. The possible amplification of heterogeneity in a formerly homogenous population makes such traits attractive for speciation theory.

One specific model which is inspired by and partially analyzed with adaptive dynamics, is presented in the next section.

### 3.4.1 Dieckmann and Doebeli, 1999

Dieckmann and Doebeli [16] proposed a computer model which links adaptive dynamics with genetic details. They demonstrated that the deterministic results in adaptive dynamics theory can be reproduced by a stochastic model. The connection is achieved by introducing a genotype-phenotype map, where the genes behave according to the known inheritance rules and the phenotype corresponds to a discretized trait.

The population has finite size, each element is equipped with a certain number of diploid, diallelic genes which all determine the phenotype, i.e., the trait value, in an additive way. Genes can take alleles $A$ and $a$. The trait value is calculated by the difference between the number of $A$-alleles and $a$-alleles. Therefore, most genotypes give rise to intermediate phenotypes. Extreme phenotypes can be produced only by genotypes which consist only of the same allele $A$ or $a$.

Mutation substitutes an $A$-allele by an $a$-allele and vice versa. As a consequence of the genotype-phenotype map where most genotypes result

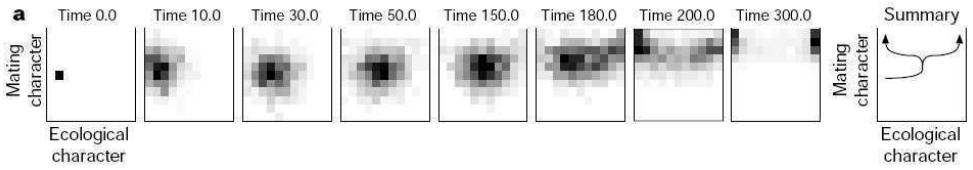



Figure 3.1: Reproduced from [16]. Population development is shown at various points in time. The darker a point, the more elements have this trait combination. Population firstly moves to an intermediate ecological trait, where it splits into two branches while assortative mating is exhibited by more and more population elements.

in intermediate phenotypes, mutation introduces a drift towards intermediate phenotypes.

Traits are linked by frequency-dependent selection. Similar traits compete with each other for limited resources and the competition decreases with the differences in trait. This means, similar individuals impact each other more strongly than dissimilar ones.

The resource function is defined as a function of the trait value. It gives the amount of resources available for a given trait. Elements with the same trait compete for these resources. The resource function allows more elements in the population with intermediate phenotypes than with extreme phenotypes. The combination of such resource function and the competition between similar phenotypes leads to a convergence stable but evolutionary unstable trait of intermediate value.

Dieckmann and Doebeli constructed the model for asexually as well as sexually reproducing populations, however, the presentation in this section is restricted to the sexual case. Two clusters of extreme pheno- and genotypes represent a polymorphism. In such a clustered population, matings between parents of distinct clusters would produce intermediate offspring and the two population branches would collapse. In order to elude this problem, Dieckmann and Doebeli implemented assortative mating. The tendency to mate assortatively is determined by a second set of diploid, diallelic genes. Possible alleles are $M$ and $m$. The mating trait is determined in additive fashion like the ecological trait. Mating traits range from assortative (only $M$-alleles) over random mating (equal number of $M$- and $m$-alleles) to disassortative mating (only $m$-alleles).

Computer simulations showed that the assortative mating trait establishes itself in the course of time and allows speciation. The speciation event was predicted by adaptive dynamics theory, because one of the intermediate ecological traits is a branching point. Once assortative mating is applied, the population branches are stabilized and do not collapse. Figure 3.1 was taken from [16] and shows the evolution in the product space of the ecological and the mating trait.

### 3.4.2 Discussion of the Dieckmann-Doebeli model

Similarity of phenotypes is a comparative property. Assortative mating requires an individual's knowledge of its own trait in order to rank mating candidates regarding their similarity and consequently their adequacy for mating. The preference for similar mates can be modified to a strategy that needs no knowledge of the own attribute. If the phenotype of preferred partner is encoded genetically for each individual, a positive correlation between own and partner's attributes might establish during evolution. If such a correlation builds up, it is equivalent to assortative mating and solves the puzzle how individuals know of their own phenotype. This model modification changes a one-allele model to a two-allele model in Felsenstein's [26] categorization. Altering the model of Dieckmann and Doebeli [16] in the described way results in equivalent simulation outcomes. The work of van Doorn and Weissing [94] investigated this model modification.

According to Dieckmann and Doebeli, assortative mating spreads when the population is concentrated at the branching point. Since the population cannot split without assortative mating, the mating genes experience an indirect selection pressure. Assortative maters which differ slightly in the trait from the evolutionary branching trait, are favored since their offspring are phenotypically situated in one of the two population branches. Offspring of random maters, in contrast, have intermediate phenotype values. However, my own reproduction (unpublished) of Dieckmann and Doebeli's model showed that the branched population is not stable, but population branches tend to collapse. An explanation might be the following: if the population is branched, the pressure for assortative mating decreases. This is because when all individuals mate assortatively, elements with a slight deviation towards less choosiness—remember that there is a mutation drift to intermediate values, that is, to random mating—should not effect the couple formation and hence the offspring formation. However, a tiny fraction of less choosy elements suffices to cause the collapse of both population branches by their cross matings. Offspring of this fraction of random maters do not experience much competition, because the rest of the population is clustered at the extreme phenotypes. These offspring have higher growth rates and the pressure for assortative mating vanishes in the population of hybrids. Once the whole population is organized again at the evolutionary branching point, it can split again. It becomes clear that the polymorphism is only stable when parameters are finely tuned and coordinated. It would be nice to have bifurcation analysis treating that issue.

The various interwoven mechanisms in Dieckmann and Doebeli's computer model, for instance the implications of the genotype-phenotype map, the resource and the competition function, and the emerging disruptive selection, complicate the clear distinction between cause and effect as well as between different causes. The large set of parameters, for instance, the competition width $\sigma_c$, parameters of the resource distribution $\sigma_K$ and

$K_0$, the birthrate $r$, the mutation rate, the offspring phenotype distribution $\sigma_{\text{off}}$, and the parameters $\sigma_d$ and $\sigma_a$ for mating probabilities make the analysis of the simulated behavior difficult.

The population experiences frequency-dependent selection which becomes disruptive at the branching point. Additionally to this disruptive selection, also assortative mating causes a splitting of the population in the direction of the extreme genotypes due to the smaller offspring genotype variance of assortative maters. This is an example where two forces work in the same direction and it would be desirable to lay open their roles separately.

Pennings et al. [70] simplified the model of Dieckmann and Doebeli by reducing the ecological trait to a single gene with three possible genotypes. The motivation was to allow for analytical treatment to support the implications of Dieckmann's and Doebeli's model. Pennings et al. proved, for instance, that a mutant with higher choosiness invades the population when both homozygotes have higher fitness values than the heterozygote.

### 3.4.3 Difference to fitness-based mating model

The Dieckmann-Doebeli model is a recent model for sympatric speciation and has received a lot of attention, probably more than all other listed models together. This is the reason why its background and details are comprehensively reviewed here.

A similarity between the Dieckmann-Doebeli model and the fitness-based mating model is that both examine the spread of a mate preference. Differences lie in the mate preferences—assortative or fitness-based mating—and the mode of examination: numerical treatment by Dieckmann and Doebeli, analytical treatment in this thesis. A further difference is the number of loci: there are several loci in the Dieckmann-Doebeli model but only two in the fitness-based mating model. Furthermore, the ecology of Dieckmann and Doebeli's model is simplified in the fitness-based mating model to two niches which are regulated separately. In contrast to the complex interactions in Dieckmann and Doebeli's model, there is no frequency-dependent selection, no resource function, and no competition based on similarity in the fitness-based model. However, in chapter 5, will present the Dieckmann-Doebeli model where I have replaced assortative mating by fitness-based mating.

## 3.5 Comparison of mate preferences

**Servedio, 2000.** Servedio [81] neatly compares two mating mechanisms for their potential of reinforcement. These two mechanisms are mate preferences for a specific trait and assortative mating.

The first locus with alleles $A$ and $a$ is used for mating decisions of females. The second locus having alleles $P$ or $p$ (and $B$ or $b$ in the assortative

model) determines the mate preference.[10] So, $P$-females prefer $A$-mates, and $p$-females prefer $a$-mates ($B$-females mate randomly, $b$-females mate assortatively with respect to the 1st-locus). Both models have haploid populations.

Both mechanisms, preferential and assortative mating, were tested for two migration scenarios: firstly, a one-way migration scheme called continental-island model where the island population receives immigrants from the continent but not vice versa. Secondly, a symmetrical migration scheme called two-island model where two populations situated on two islands symmetrically exchange migrants. Viability selection acts on the first locus only and is favoring $A$ in one geographical spot and $a$ in the other. Hybrids have lower viability fitness due to two epistatically interacting background loci.

In the beginning, all individuals carry the alleles $P$ (or $B$) and all, also $P$-carriers, mate randomly. The population finds itself in a selection-migration equilibrium before the mating alleles ($p$ or $b$) are introduced which cause non-random mating. In the preference model, $P$ no longer causes random mating, instead, $P$-carriers prefer $A$-mates and $p$-carriers prefer $a$-mates. If $p$ (or $b$) spreads, the condition for reinforcement is fulfilled and a premating isolation mechanism is established.

Using simulations, Servedio found that reinforcement occurs under a larger parameter range in the assortative two-island model than in the preference two-island model. Considering the continental-island models, the preference model leads more often to reinforcement than the assortative continental-island model.

Servedio points out that it can hardly be determined in practice whether preference or assortative mating prevails. Because under preference mating, the mating allele $P$ will get linked to the preferred trait allele $A$, and $p$ will get linked to $a$. So, $PA$-carriers mate among themselves and $pa$-carriers mate among themselves. Thus, the mating behavior cannot be told apart from assortative mating—where each group of $A$-carriers and $a$-carriers mates among itself—by mere observation. This is one way where different genetic mechanisms can result in the same mating behavior.

### 3.5.1 Difference to fitness-based mating model

Servedio [81] and chapter 5 of my thesis both compare assortative mating vs. preferential mating under comparable conditions[11] with numerical simulations. However, Servedio's preference model is a two-allele model in Felsenstein's categorization [26] whereas the assortative mating model is a one-allele model. Furthermore, the focus in [81] lies on the potential for

[10] The original naming is $T_1$ and $T_2$ as the 1st-locus alleles and $P_1$ and $P_2$ (or $B_1$ and $B_2$ in the assortative model) as the 2nd-locus alleles.

[11] It is not clear whether choosiness implies a cost for females in Servedio's model. The author refers for model details to the previous publication [82] in which each female mates once and thus, there are no costs. It is unclear whether the model modifications in [81] leave that detail unchanged.

reinforcement. In my thesis, both mate preferences have been contrasted in terms of their biological justification (section 2.4) and will be compared in terms of their evolutionary justification (section 4.2.4) and their transient behavior as well as their impact on the genetic composition (chapter 5).

The preference two-island model of Servedio [82] differs to the fitness-based mating-model in two aspects. Firstly, the $P$ and $p$-alleles cause a fixed preference for $A$ and $a$-carriers indifferently of the niche (island). Fitness-based maters, in contrast, prefer the fittest of their niche irrespective of their phenotype. In fact, a female preferring say $A$-mates in one niche prefers $a$-mates after the migration to the other niche. Secondly, fitness-based mating competes with random mating whereas allele $P$ produces random mating behavior only when the $p$-allele has not yet been introduced in the population.

## 3.6 Evolution of mate preferences

**Matessi et al., 2001.** In an extremely well written paper, Matessi, Gimelfarb, and Gavrilets [57] examine the effect of disruptive selection on the evolution of assortative mating as a premating isolation mechanism.

Matessi et al. showed that disruptive selection per se is not sufficient to bring a population from incipient assortative mating to reproductive isolation. Any level of disruptive selection, however, is sufficient to enable invasion of weak assortment in randomly mating population which is in polymorphic equilibrium, and prevent invasion of lower assortment in reproductively isolated subpopulations. Driving a population throughout from low levels to high levels of assortment is conditioned on sufficiently strong disruptive selection.

### 3.6.1 Relevance for fitness-based mating model

Studying mate preferences includes not only their effect on the population composition but includes also their evolutionary justification. Matessi et al. provide an elucidating analysis of disruptive selection as a prerequisite for spread and fixation of assortative mating. In the fitness-based mating model no such exterior cause of selection pressure towards a mate preference needs to be introduced. Instead, the existence of selection pressure can be observed by the spread and fixation of the mate preference. The preference of fit mates over others is, however, easier to justify within the evolutionary framework than assortative mating. Whereas the benefit from a fit partner is obvious, the benefit from a like partner needs further assumptions. For instance, through prevailing disruptive selection, which can favor offspring of like partners. Knowing that, Matessi et al.'s contribution is insightful and similar work is desired for fitness-based mating once hidden assumptions for selection pressure in favor of fitness-based mating are uncovered.

## 3.7 Classification of the fitness-based mating model

The model to be presented in this thesis investigates how female choosiness based on direct advantages can cause and maintain a polymorphic population. The fitness-based mating model is a 2-locus model with two niches, where the population size of each niche is regulated separately. The first locus is subject to natural selection and the second locus determines the strength of the preference for fit mates. Fitness-based mating is a preference-model. The preferred trait is not fixed instead, the preferred trait is flexible, i.e., it depends on whatever geno-/phenotype performs best in its ecological setting. Therefore, the preferred trait can vary across the niches. The fitness-based mate preference competes with random mating and does not imply costs of choosiness. The population is separated into females and males, among which only females exhibit the mate preference. Ecological selection acts on both sexes alike.

# 4 Fitness-based mating model

The first section of this chapter is devoted to the step by step introduction of fitness-based mating. I will use toy models, give extensive explanations, and I will discuss in detail the derivation of the mating probabilities. At first, I will look at the couple formation process itself and then introduce fitness-based mating in the haploid setting.

The second section of this chapter treats the model dynamics analytically. The focus will be on the conditions for the spread of fitness-based mating and on the population composition.

## 4.1 The stepwise introduction of fitness-based mating

### 4.1.1 Drawing couples

At the beginning, individuals are assumed to be hermaphrodites, meaning they possess female as well as male germ cells and are able to fulfill both sexual roles: female as well as male. Consequently, two arbitrary individuals can mate with each other. Later, genders will be introduced for modelling different degrees of choosiness.

I will abstractly describe the couple formation in a population as drawing balls from an urn. Each individual is represented by a ball and has the same probability of being drawn as any other ball/individual. Two partners which will mate correspond to drawing two balls at the same time from the urn. Mating schemes such as preferences are realized by accepting or rejecting certain mating couples. Rejected mating couples are placed back into the urn. Accepted couples, in contrast, mate and form offspring. Consequently, the probability of drawing two partners $i, j$ denoted by $P_{\text{meet}}(i, j)$ is distinguished from their probability of mating $P_{\text{mate}}(i, j)$.[1]

[1] There exists a large variety of notations on this matter. For example, $P_{\text{meet}}$ and $P_{\text{mate}}$ correspond to $P_i P_j$ and $X_{i,j}$ in [65, 63] or to $\phi_{f,i}\phi_{m,j}$ and $P(i \times j)$ in [29]. The list is not complete and the reader may excuse the introduction of yet another notation. I hope, the reader will soon be convinced by its intuitive character. Often, populations are modelled as infinitely large to avoid finite size effects and to approximate probabilities by relative frequencies. For this reason, I will label probabilities as well as relative frequencies with "P", because I assume that the latter approximates the former. In doing so, basic probabilistic assumptions like the sum of probabilities over the whole sample space resulting in 1 do not have to be listed explicitly. I will label probabilities with explanatory indices, so that formulas allow for an intuitive understanding.

### Mating systems: Monogamy or polygamy

After mating, a couple can be placed back into the urn, so that each of them has a chance to be redrawn in a new mating couple.

Mating systems are, for instance, monogamy or polygamy. Polygamy corresponds to placing couples back into the urn after mating. Monogamy corresponds to their removal from the urn. Individuals can mate more than once under polygamy and mate at most once under monogamy. Both schemes, monogamy and polygamy, mark the most common mating schemes. Other schemes, like putting only one partner back into the urn—possibly depending on its gender—or allowing a specific number of matings, can be modelled as well, but we can concentrate on the mentioned ones for now.

### No removal from urn - Polygamy

This paragraph contains the analysis of drawing mating couples under polygamy. Mating couples will be placed back into the urn after mating and thus the urn content of balls does not change during the process. The advantage of polygamy is that the probability for drawing a couple does neither depend on the number nor on the composition of formerly drawn couples. Consider an urn with balls of two types: $A$ and $B$. The number of $A$-balls is $\alpha$ and the number of $B$-balls is $\beta$.

**Random mating.** First, I will look at random mating. Later, I will highlight the changes that will occur when a mate preference is introduced. The probability of drawing a certain couple $AA$, $AB$, or $BB$ is for each drawing

$$P_{\text{meet}}(AA) = \begin{cases} \frac{\binom{\alpha}{2}}{\binom{\alpha+\beta}{2}} = \frac{\alpha(\alpha-1)}{(\alpha+\beta)(\alpha+\beta-1)}, & \text{if } \alpha+\beta \geq 2 \\ 0, & \text{otherwise,} \end{cases} \tag{4.1}$$

$$P_{\text{meet}}(AB) = \frac{\binom{\alpha}{1}\binom{\beta}{1}}{\binom{\alpha+\beta}{2}} = \frac{2\alpha\beta}{(\alpha+\beta)(\alpha+\beta-1)}, \tag{4.2}$$

$$P_{\text{meet}}(BB) = \begin{cases} \frac{\binom{\beta}{2}}{\binom{\alpha+\beta}{2}} = \frac{\beta(\beta-1)}{(\alpha+\beta)(\alpha+\beta-1)}, & \text{if } \alpha+\beta \geq 2 \\ 0, & \text{otherwise.} \end{cases} \tag{4.3}$$

Like most authors I will assume that $\alpha$ and $\beta$ are so large that that the approximations $\alpha \approx \alpha - 1$ and $\beta \approx \beta - 1$ are valid. In this way, the drawing probabilities simplify to

$$P_{\text{meet}}(AA) = \frac{\alpha^2}{(\alpha+\beta)^2} \tag{4.4}$$

$$P_{\text{meet}}(AB) = \frac{2\alpha\beta}{(\alpha+\beta)^2} \tag{4.5}$$

$$P_{\text{meet}}(BB) = \frac{\beta^2}{(\alpha+\beta)^2}. \tag{4.6}$$

Individuals mate randomly, they do not show preferences and accept any partner. Hence, meeting probabilities are the same as mating probabilities: $P_{\text{meet}} = P_{\text{mate}}$.

When does the drawing of balls stop? This question is equivalent to "How many couples can be formed or how many matings can at most take place?" A fixed number of drawings induces finite-size effects. Therefore, one usually assumes that infinitely many drawings are performed such that the fractions of the couples $AA$, $AB$, $BB$ among all drawn couples is the same as $P_{\text{meet}}(AA)$, $P_{\text{meet}}(AB)$, $P_{\text{meet}}(BB)$. The law of large numbers justifies that assumption, because it says that the empirical average of a random variable converges to the mean of the random variable's distribution.

**Mate preference.** Now I will introduce a mate preference. Consider that $A$-balls accept matings with $B$-balls and reject $A$-balls. In contrast, $B$-balls have no mate preference. If an $AA$-couple is drawn, it will instantly be placed back into the urn. Therefore, only $AB$- and $BB$-couples mate. Mating couples are put back into the urn after mating. In contrast to random mating, the meeting and mating probabilities are no longer the same.

The mate preference is modelled by introducing an acceptance probability $P_{x_i}^{\text{accept}}(x_j)$ with which the ball $x_i$ accepts a mating with $x_j$. Since $A$-balls reject other $A$-balls, and $B$-balls accept all, $P^{\text{accept}}$ is set to:

$$P_A^{\text{accept}}(A) = 0, \quad P_A^{\text{accept}}(B) = P_B^{\text{accept}}(A) = P_B^{\text{accept}}(B) = 1. \tag{4.7}$$

The mating probability is then set to

$$P_{\text{mate}}(x_i, x_j) = \frac{1}{N_P} P_{\text{meet}}(x_i, x_j) P_{x_i}^{\text{accept}}(x_j) P_{x_j}^{\text{accept}}(x_i) \tag{4.8}$$

with

$$N_P = \sum_{(x_i, x_j)} P_{\text{meet}}(x_i, x_j) P_{x_i}^{\text{accept}}(x_j) P_{x_j}^{\text{accept}}(x_i) \tag{4.9}$$

$$= P_{\text{meet}}(AB) + P_{\text{meet}}(BB), \tag{4.10}$$

where the sum in equation (4.9) runs through index pairs, not over all combinations of two individuals. This means that $(x_i, x_j) = (x_j, x_i)$ and that a couple enters the sum only once.

The set of mating couples is a subset of all couples that have met, because meeting is a precondition for mating. Most publications concentrate on the set of mating couples and thus calculate the fraction of specific couples among those that have mated. This is justified by assuming constant population sizes, when fewer couples reproduce the more and thus compensate for couples that do not mate. That is a realistic approach, firstly, because all living creatures have the potential to produce more offspring than the number of parents [56] and secondly, because stress, population density, and general intra-specific competition reduce reproduction rates

(e.g., aphids [43] or sparrows [3]). Therefore, the mating probability is normalized by $N_P$. Thus,

$$P_{\text{mate}}(AA) = 0 \tag{4.11}$$

$$P_{\text{mate}}(AB) = \frac{P_{\text{meet}}(AB)}{P_{\text{meet}}(AB) + P_{\text{meet}}(BB)} \tag{4.12}$$

$$P_{\text{mate}}(BB) = \frac{P_{\text{meet}}(BB)}{P_{\text{meet}}(AB) + P_{\text{meet}}(BB)}. \tag{4.13}$$

Returning to the urn model, the probability of mating for a specific pair is then

$$P_{\text{mate}}(AA) = 0 \tag{4.14}$$

$$P_{\text{mate}}(AB) = \begin{cases} \frac{\binom{\alpha}{1}\binom{\beta}{1}}{\binom{\alpha}{1}\binom{\beta}{1}+\binom{\beta}{2}} = \frac{2\alpha}{2\alpha+\beta-1}, & \text{if } \alpha, \beta \geq 1 \text{ or } \beta > 1 \\ 0, & \text{otherwise,} \end{cases} \tag{4.15}$$

$$P_{\text{mate}}(BB) = \begin{cases} \frac{\binom{\beta}{2}}{\binom{\alpha}{1}\binom{\beta}{1}+\binom{\beta}{2}} = \frac{\beta-1}{2\alpha+\beta-1}, & \text{if } \alpha, \beta \geq 1 \text{ or } \beta > 1 \\ 0, & \text{otherwise.} \end{cases} \tag{4.16}$$

When we assume that $\alpha$ and $\beta$ are large enough, i.e., $\beta - 1 \approx \beta$, it simplifies the mating probabilities to:

$$P_{\text{mate}}(AA) = 0 \tag{4.17}$$

$$P_{\text{mate}}(AB) = \frac{2\alpha}{2\alpha + \beta} \tag{4.18}$$

$$P_{\text{mate}}(BB) = \frac{\beta}{2\alpha + \beta}. \tag{4.19}$$

#### Removal from urn - Monogamy

If the population is monogamous, couples that have mated are removed from the urn. Therefore, the number of $A$ and $B$-balls changes over time and likewise do the meeting probabilities. Therefore, we need to monitor how many balls are left in the urn. The best way to do this is counting the number $AA$-, $AB$-, and $BB$-couples, from which the remaining number of balls can be inferred. Let $l$ count $AA$-couples, let $m$ count $AB$-couples, and let $n$ count $BB$-couples.

**Random mating.** As before, two balls are drawn at a time. Thereafter, they are removed from the urn and the vector of $AA$-, $AB$-, and $BB$-pairs, $(l, m, n)$, is updated. In such procedure, $(l, m, n)$ updates to $(l + 1, m, n)$, $(l, m + 1, n)$, or $(l, m, n + 1)$ when an $AA$-pair, $AB$-pair, or a $BB$-pair has been drawn. The vector $(l, m, n)$ remains unchanged when there are less than two balls left in the urn. The number of remaining $A$- and $B$-balls, denoted by $a$ and $b$ can be inferred from $(l, m, n)$:

$$a = \alpha - 2l - m, \quad b = \beta - 2n - m. \tag{4.20}$$

Thus, the probabilities of drawing a certain pair given $(l, m, n)$ are:

$$P_{\text{meet}}(AA|l,m,n) = \begin{cases} \frac{\binom{a}{2}}{\binom{a+b}{2}} = \frac{a(a-1)}{(a+b)(a+b-1)}, & \text{if } a+b \geq 2 \\ 0, & \text{otherwise} \end{cases} \tag{4.21}$$

$$P_{\text{meet}}(AB|l,m,n) = \frac{\binom{a}{1}\binom{b}{1}}{\binom{a+b}{2}} = \frac{2ab}{(a+b)(a+b-1)}, \tag{4.22}$$

$$P_{\text{meet}}(BB|l,m,n) = \begin{cases} \frac{\binom{b}{2}}{\binom{a+b}{2}} = \frac{b(b-1)}{(a+b)(a+b-1)}, & \text{if } a+b \geq 2 \\ 0, & \text{otherwise.} \end{cases} \tag{4.23}$$

We find the relative frequency of a specific configuration $(l, m, n)$ among all possible configurations, denoted by $P(l, m, n)$, when we look at the sequence of single balls. Each sequence contains $\alpha$ $A$-balls and $\beta$ $B$-balls and has the same probability namely, 1 over the number of ways to distribute $\alpha$ balls among $\alpha + \beta$ positions: $1/\binom{\alpha+\beta}{\alpha}$. The probability of the configuration $(l, m, n)$ is the product of the sequence probability and the number of sequences that lead to the configuration. Let us group the sequence into pairs by starting at the first position. If $\alpha + \beta$ is odd, the last ball is not part of a pair, it is left in the urn. The number of pairs in each sequence is $N = \lfloor(\alpha + \beta)/2\rfloor$. There are $\binom{N}{l,m,n}$ possibilities to distribute $l$ $AA$, $m$ $AB$, and $n$ $BB$-pairs among the $N$ positions. Since an $AB$-pair is not distinguished from a $BA$-pair, an additional factor of 2 is multiplied to $\binom{N}{l,m,n}$ for each mixed pair. Thus, the probability of $(l, m, n)$ is

$$P(l,m,n) = \frac{2^n}{\binom{\alpha+\beta}{\alpha}}\binom{N}{l,m,n}. \tag{4.24}$$

Please note that the number of drawn $AB$-pairs, $m$, determines the number of $AA$ and $BB$-pairs, because the remaining $A$-balls, that are $\alpha - m$, form $AA$-couples, and the remaining $B$-balls, that are $\beta - m$, form $BB$-couples.

The meeting probability of a couple changes during the process of drawing couples. Therefore, I will use the fraction of a couple among all mated couples to approximate its mating probability. The expected number of $AA$-pairs, $AB$-pairs, and $BB$-pairs will be denoted by $\phi_{AA}$, $\phi_{AB}$, and $\phi_{BB}$, and then I set:

$$P_{\text{mate}}(AA) = \frac{\phi_{AA}}{N}, \quad P_{\text{mate}}(AB) = \frac{\phi_{AB}}{N}, \quad P_{\text{mate}}(BB) = \frac{\phi_{BB}}{N}. \tag{4.25}$$

Now we want to know what is the expected relative frequency of a couple among all possible drawings. Let us label the balls in the urn by the numbers $1, 2, \ldots, \alpha + \beta$. The set of all possible drawings is the set of all permutations of these numbers. Let us now group the numbers of an arbitrary sequence, say, $i_1, i_2, \ldots$ into pairs $(i_1, i_2)$, $(i_3, i_4), \ldots$. If $\alpha + \beta$ is odd, a single number/ball will remain, but this will not change the argument. I will introduce an equivalence relation: Two pair-sequences

are equivalent if the one is a pair-permutation of the other. A pair-permutation of $\ldots,(i_k,i_{k+1}),\ldots,(i_l,i_{l+1}),\ldots$ is either an exchange of balls within a pair, i.e., changing $(i_k,i_{k+1})$ to $(i_{k+1},i_k)$, or an exchange of pairs, i.e. $\ldots,(i_k,i_{k+1}),\ldots,\ (i_l,i_{l+1}),\ldots$ to $\ldots,(i_l,i_{l+1}),\ldots,(i_k,i_{k+1}),\ldots$, or any combination of such within-pair exchanges or pair-exchanges. Using this equivalence relation, we can group the set of pair-sequences into equivalence classes where the numbers of $AA$, $AB$, or $BB$-pairs are the same in each pair-sequence of one equivalence class.

Look now at an arbitrary pair-sequence in an arbitrary equivalence class. The probability of drawing a pair $AA$, $AB$, or $BB$ out of the pairs of this pair-sequence is the same as its relative frequency in this pair-sequence. This is valid for all pair-sequences in this class, because the number of $AA$, $AB$, and $BB$-pairs is the same in all pair-sequences. Furthermore, the same reasoning holds in each equivalence class. Consequently, we know that the expected relative frequency of a pair $AA$, $AB$, or $BB$ after the drawing process must be the same as its frequency in the urn with $\alpha$ $A$-balls and $\beta$ $B$-balls:

$$\frac{\phi_{AA}}{N}=\frac{\binom{\alpha}{2}}{\binom{\alpha+\beta}{2}},\quad \frac{\phi_{AB}}{N}=\frac{\binom{\alpha}{1}\binom{\beta}{1}}{\binom{\alpha+\beta}{2}},\quad \frac{\phi_{BB}}{N}=\frac{\binom{\beta}{2}}{\binom{\alpha+\beta}{2}}. \tag{4.26}$$

**Mate preference.** Now again, $A$ refuses to mate with another $A$. Thus, $AA$-couples are placed instantly back into the urn, whereas $AB$- and $BB$-couples mate and are removed. The numbers of $A$ and $B$-balls do not only shrink, but also the ratio of $A$ and $B$-balls changes. This time, we count only $AB$- and $BB$-couples, the vector $(m,n)$ sufficiently describes the outcome.

Let $P_{\text{meet}}(AB|m,n)$ and $P_{\text{meet}}(BB|m,n)$ be the probability that an $AB$-pair, or $BB$-pair is drawn under the condition that already $m$ $AB$- and $n$ $BB$-pairs have been drawn. Drawing a further $AB$-pair increases $m$ by 1 but leaves $n$ unchanged. Analogously, drawing a $BB$-pair increases $n$ by 1 and leaves $m$ unchanged. The drawing probability depends on the numbers of $A$- and $B$-balls left in the urn. These numbers can be inferred by $m$ and $n$. If $a$ and $b$ denote again the number of remaining $A$- and $B$-balls, it holds that

$$a=\alpha-m,\quad b=\beta-2n-m. \tag{4.27}$$

The acceptance probability $P^{\text{accept}}$ is as it has been defined in equation (4.7). Mating probabilities are normalized by the fraction of all mating couples $N_P$. Similar to the equations (4.14)–(4.16), the conditional mating

probabilities $P_{\text{mate}}(AB|m,n)$ and $P_{\text{mate}}(BB|m,n)$ are

$$P_{\text{mate}}(AA|m,n) = 0 \tag{4.28}$$

$$P_{\text{mate}}(AB|m,n) = \begin{cases} \frac{\binom{a}{1}\binom{b}{1}}{\binom{a}{1}\binom{b}{1}+\binom{b}{2}} = \frac{2a}{2a+b-1}, & \text{if } a,b \geq 1 \text{ or } b>1 \\ 0, & \text{otherwise,} \end{cases} \tag{4.29}$$

$$P_{\text{mate}}(BB|m,n) = \begin{cases} \frac{\binom{b}{2}}{\binom{a}{1}\binom{b}{1}+\binom{b}{2}} = \frac{b-1}{2a+b-1}, & \text{if } a,b \geq 1 \text{ or } b>1 \\ 0, & \text{otherwise.} \end{cases} \tag{4.30}$$

If the final distribution $P(m,n)$ of the number of $AB$- and $BB$-couples drawn is known, their expected numbers can be calculated and then also their mating probabilities with the help of equation (4.25). But usually, the derivation of these mating probabilities is tedious and they are avoided by either ignoring the changing fraction $A/B$-balls and/or by assuming polygamy.

The difficulty of deriving an explicit formula for $P(m,n)$ lies in the non-exchangeability of the process. That means that the specific drawing sequences that lead to $(m,n)$ might have—and usually do have—different probabilities such that the sum of those is enormously long. Even the construction of a Markov chain for the process does not bring us any further, because the transition matrix is not irreducible and cannot be subdivided into irreducible parts. The Perron-Frobenius-Theorem is therefore not applicable. However, it is possible to derive an iteration formula for the expected number of a mating pair and then also for their mating probabilities. Let us again first look at the case without mate preferences. Let $\phi_{BB}(\alpha,\beta)$ be the expected number of mated $BB$-pairs when the initial numbers of $A$ and $B$-balls are $\alpha,\beta$. We then have the recursion formula

$$\phi_{BB}(\alpha,\beta) = \frac{\binom{\beta}{2}}{\binom{\beta+\alpha}{2}}\big(1+\phi_{BB}(\alpha,\beta-2)\big)+\frac{\beta\alpha}{\binom{\beta+\alpha}{2}}\phi_{BB}(\alpha-1,\beta-1)+\frac{\binom{\alpha}{2}}{\binom{\beta+\alpha}{2}}\phi_{BB}(\alpha-2,\beta). \tag{4.31}$$

The first factor in each summand tells us the probability of the corresponding first move (two $B$-balls, etc.), whereas the second one yields the expectation values for the situation after the first move, that is, when two balls have been taken out of the urn.

We have a similar formula in the case of mate preferences, denoting the expected number of mated $BB$-pairs now by $f_{BB}(\alpha,\beta)$.

$$f_{BB}(\alpha,\beta) = \frac{\binom{\beta}{2}}{\binom{\beta+\alpha}{2}}\big(1+f_{BB}(\alpha,\beta-2)\big)+\frac{\beta\alpha}{\binom{\beta+\alpha}{2}}f_{BB}(\alpha-1,\beta-1)+\frac{\binom{\alpha}{2}}{\binom{\beta+\alpha}{2}}f_{BB}(\alpha,\beta). \tag{4.32}$$

The crucial difference to equation (4.31) is that in the very last term we now have $\alpha$ instead of $\alpha-2$, because two $A$-balls are returned to the urn.

From the equations (4.32) and (4.29)–(4.30), we obtain

$$f_{BB}(\alpha,\beta) = \frac{1}{\binom{\beta}{2}+\beta\alpha}\left[\binom{\beta}{2}(1+f_{BB}(\alpha,\beta-2)) + \beta\alpha f_{BB}(\alpha-1,\beta-1)\right]. \tag{4.33}$$

Analogously, we obtain

$$f_{AB}(\alpha,\beta) = \frac{1}{\binom{\beta}{2}+\beta\alpha}\left[\binom{\beta}{2} f_{AB}(\alpha,\beta-2) + \beta\alpha(1+f_{AB}(\alpha-1,\beta-1))\right]. \tag{4.34}$$

We also observe that

$$2f_{BB}(\alpha,\beta) + f_{AB}(\alpha,\beta) = \beta \text{ for } \alpha \geq \beta \text{ and } \beta-\alpha \text{ even}. \tag{4.35}$$

Thus, in this case, the formula for $f_{AB}$ directly follows from the one for $f_{BB}$. One can derive the following explicit expressions:

$$f_{AB}(1,\beta) = \begin{cases} \frac{\beta}{\beta+1}, & \text{if } \beta \text{ even} \\ 1, & \text{if } \beta \text{ odd} \end{cases} \tag{4.36}$$

$$f_{AB}(2,\beta) = \begin{cases} \frac{2\beta(\beta+4)}{(\beta+1)(\beta+3)}, & \text{if } \beta \text{ even} \\ \frac{2}{(\beta+1)(\beta+3)}\left(6+(\beta-1)(\beta+4)-2\sum_{k=0}^{n-1}\frac{1}{2k+1}\right), & \text{if } \beta = 2n+1. \end{cases} \tag{4.37}$$

and in general, we have

$$f_{BB}(\alpha,0) = 0 \tag{4.38}$$

$$f_{BB}(\alpha,1) = 0 \tag{4.39}$$

$$f_{BB}(\alpha,2) = \frac{1}{1+2\alpha} \tag{4.40}$$

$$f_{BB}(\alpha,3) = \frac{3\alpha-1}{2\alpha^2+\alpha-1} \tag{4.41}$$

$$f_{BB}(\alpha,4) = \frac{24\alpha^2-16\alpha-26}{8\alpha^3+4\alpha^2-18\alpha-9}, \tag{4.42}$$

and further such formulae can be derived by iteration. However, it would be nice to have a closed formula.

### 4.1.2 Haploid model with two loci

This section covers balls with two features instead of one. Balls have now either the color red or blue and are either speckled or not. Altogether there are now four types of balls in the urn: red/speckled, red/not-speckled, blue/speckled, and blue/not-speckled. The urn model can be translated into biological terms: Each ball corresponds to a single individual, each

| $P^{\text{accept}}_{x_j}(x_i)$ | $x_j$ | | | |
|---|---|---|---|---|
| $x_i$ | $AM$ | $Am$ | $aM$ | $am$ |
| $AM$ | 1 | 1 | 1 | 1 |
| $Am$ | 1 | 1 | 1 | 1 |
| $aM$ | 0 | 1 | 0 | 1 |
| $am$ | 0 | 1 | 0 | 1 |

Table 4.1: The acceptance probabilities

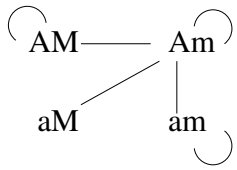



Figure 4.1: The mate preference. The gametes are vertices and if the mate preferences of two gametes are met, an edge is drawn between them.

feature (color, pattern) to a locus in the genome, and each feature value (red/blue, speckled/not-speckled) to an allele.

Consider two loci with the 1st-locus alleles $A$ (red) and $a$ (blue), and the 2nd-locus alleles $M$ (speckled) and $m$ (not-speckled). I will concentrate on haploid populations, this means I will look at gametes. There are four different gametes: $AM$, $Am$, $aM$, and $am$. For the corresponding urn model it means that the $A$- and $B$-balls are replaced by $AM$-, $Am$-, $aM$-, and $am$-balls.

The second locus determines whether an individual mates randomly or is choosy about a mate. If an individual with an $m$-allele meets an arbitrary other individual $j$, it accepts mating:

$$P^{\text{accept}}_{Am}(j) = P^{\text{accept}}_{am}(j) = 1. \tag{4.43}$$

Individuals with an $M$-allele, however, reject $a$-carriers and accept only $A$-carriers:

$$\begin{aligned} &P^{\text{accept}}_{AM}(AM) = P^{\text{accept}}_{AM}(Am) = P^{\text{accept}}_{aM}(AM) = P^{\text{accept}}_{aM}(Am) = 1 \\ &P^{\text{accept}}_{AM}(aM) = P^{\text{accept}}_{AM}(am) = P^{\text{accept}}_{aM}(aM) = P^{\text{accept}}_{aM}(am) = 0. \end{aligned} \tag{4.44}$$

Table 4.1 lists the acceptance probabilities. The two loci are connected in the sense that the 2nd-locus allele influences the mating decision on grounds of the mate's 1st-locus allele. Figure 4.1 visualizes the mating behaviour by means of a graph.

$M$-individuals reject mates with an $a$-allele, so there are only the matings between the balls given in table 4.2's first column. Let $p(AM)$, $p(Am)$, $p(aM)$, and $p(am)$ denote the relative gamete frequencies. Gametes meet with the probabilities given in the second column of table 4.2. The mating

| couple | $P_{\text{mate}} \cdot N_P$ |
|---|---|
| $AM, AM$ | $p(AM)^2$ |
| $AM, Am$ | $2p(AM)p(Am)$ |
| $Am, Am$ | $p(Am)^2$ |
| $Am, am$ | $2p(Am)p(am)$ |
| $Am, aM$ | $2p(Am)p(aM)$ |
| $am, am$ | $p(am)^2$ |

Table 4.2: Mating probabilities in the haploid model.

probabilities, which we can calculate with the help of equation (4.8), are normalized by the fraction $N_P$ of all mating couples:

$$N_P = \Big(p(AM) + p(Am)\Big)^2 + p(am)^2 + 2p(Am)\Big(p(aM) + p(am)\Big). \quad (4.45)$$

Let $p(M)$, $p(m)$, $p(A)$, and $p(a)$ denote the relative allele frequencies, so it also holds that

$$N_P = p(A)^2 + p(am)^2 + 2p(Am)p(a) \quad (4.46)$$

and

$$N_P = p(AM)^2 + p(am)^2 + p(Am)\Big(2 - p(Am)\Big). \quad (4.47)$$

We would like to know which of the alleles is more successful: $M$ or $m$. An $M$-carrier has less mating opportunities than an $m$-carrier, because it rejects a fraction of potential mates. $M$- and $m$-individuals have therefore different mating rates. While $m$-balls will mate with whatever type of partner, hence have a mating rate of 100%, $M$-balls are restricted to $A$-partners; their mating rate is limited by the fraction of $A$-balls in the urn. Thus, $M$-balls are disadvantaged by their mate preference in contrast to randomly mating individuals. Therefore, choosiness is costly. I want to quantify the mating disadvantage of $M$-carriers.[2] In that matter, we can ignore matings between gametes of different 2nd-locus alleles, because these matings do not change the ratio of $M/m$-alleles. Hence, I will look only at matings where the both 2nd-locus alleles are the same. These are: $AM$ and $AM$, $Am$ and $Am$, $Am$ and $am$, and $am$ and $am$. Let $N_F$ denote the sum of all these matings. Among these matings, only $AM$ and $AM$, which have a share of

$$\frac{p(AM)^2}{N_P N_F} \quad (4.48)$$

[2] Another concept is the relative mating success of genotype $x_i$ defined by

$$\frac{\sum_{(x_i,x_j)} P_{\text{mate}}(x_i, x_j)}{p(x_i)}$$

in [29], p. 281. It will be used in section 4.2.3, but for the question now, it will give no explanatory hints.

produce $M$-offspring. The share of matings that account for $m$-offspring is

$$\frac{p(Am)^2 + 2p(Am)p(am) + p(am)^2}{N_P N_F} = \frac{p(m)^2}{N_P N_F}. \tag{4.49}$$

Thus, the ratio $p(AM)/p(m)$ determines which 2nd-locus allele will dominate in the subsequent generation: If $p(AM) \geq p(m)$, then $p(M) \geq p(m)$ will hold in the next generation. However, this relationship only holds in two subsequent generations. It does not indicate which $M/m$-allele will dominate the population in the end.

It is clear, and can be confirmed by straightforward computations, that the relative frequency of $aM$ always decreases from one generation to the next, because this type finds the fewest mating partners. Likewise, the relative frequency of $Am$ increases in most situations when both $AM$ and $am$ are present in the population. This is because the $Am$-type finds more mating partners than the others. Therefore, the dynamics will converge to a situation where $aM$ and also either $AM$ or $am$ is extinct. This is a situation where either the mating preference as encoded by $M$ has disappeared from the population or it has become irrelevant because only the preferred mating type $A$ survives.

So, let $p'(aM)$ denote the frequency of $aM$ in the subsequent generation, then

$$p'(aM) = \frac{1}{4} P_{\text{mate}}(aM, Am) \overset{\text{Table 4.2}}{=} \frac{1}{2N_P} p(aM)p(Am). \tag{4.50}$$

Since $N_P > \frac{1}{2}p(Am)$ which is—due to equation (4.47)—equivalent to

$$p(AM)^2 + p(am)^2 + p(Am)\Big(\frac{3}{2} - p(Am)\Big) > 0 \tag{4.51}$$

and is always true, it holds that $p'(aM) < p(aM)$. Thus, the relative frequency of $aM$ shrinks from one generation to the next and the equilibrium[3] population will lack any gamete $aM$.

Similarly,

$$p'(Am) = \frac{1}{N_P}\left(p(Am)^2 + p(AM)p(Am) + p(Am)p(am) + \frac{1}{2}p(aM)p(Am)\right) \tag{4.52}$$

$$= \frac{p(Am)}{N_P}\left(1 - \frac{1}{2}p(aM)\right). \tag{4.53}$$

And $p'(Am) > p(Am)$ is equivalent to

$$2p(AM)p(am) + p(aM)\left(\frac{1}{2} - p(Am) + p(AM) + p(am)\right) > 0. \tag{4.54}$$

[3]The terms *fixed point* and *equilibrium* are used exchangeably in this thesis. In equilibrium, each gametic frequency remains constant, i.e., it has reached a fixed point. Since a population can be described by a vector whose components are the gametic frequencies, I will say that also the population has reached a fixed point.

This condition is mostly fulfilled except when the relative frequency of $Am$ is high and those of the other gametes is relatively low. In these cases, even though $Am$ has the most mating partners, its matings with $aM$ lead to a decrease of $p(Am)$. This is because three-quarters of the offspring produced by the matings between $Am$ and $aM$ have genotypes other than $Am$, and only one-quarter has genotype $Am$. Thus, as long as equation (4.54) is not fulfilled, $p(Am)$ decreases. However, these situations do not persist for many generations. Shortly after, $Am$ profits from the abundance of potential mates. As claimed, $p(Am)$ increases on the long run, but in any case—and this will be proofed below—it increases as long as both types $AM$ and $am$ are present in the population and $p(aM) = 0$.

In conclusion, $aM$ and at least one of $AM$ and $am$ will become extinct asymptotically. As there will then be no mating differences between the remaining types ($AM$ and $Am$ or $am$ and $Am$), any distribution of those two types is possible as an asymptotic limit, and which distribution will be realized will depend on the initial conditions. The only exception occurs when both $Am$ and $aM$ are already initially absent. Since $AM$ and $am$ do not mate with each other, the one that is initially more frequent will gain the upper hand. An equilibrium between these types then is only possible when they both exist with the same relative frequency $\frac{1}{2}$, and this then is an unstable equilibrium.

**Fixed point analysis.** Naming the gametes $x_1, x_2, x_3$, and $x_4$ for $AM, Am, aM$, and $am$, respectively, the subsequent relative gametic frequencies $x_i'$ can be calculated by

$$x_i' = \sum_{(x_j,x_k)} P_{\text{mate}}(x_j, x_k) P_{\text{off}}(x_i | x_j, x_k) \tag{4.55}$$

where $P_{\text{mate}}$ can be obtained from table 4.2 where $N_P$ is given in equation (4.45), and $P_{\text{off}}$ is given in equation (2.3). The sum goes over all couples.

The following system of recurrence equations gives the population development from one generation to the next:

$$\begin{aligned}
x_1' &= \frac{1}{N_P}(x_1^2 + x_1x_2 + \frac{1}{2}x_2x_3) && (4.56)\\
x_2' &= \frac{1}{N_P}(x_2^2 + x_1x_2 + x_2x_4 + \frac{1}{2}x_2x_3) && (4.57)\\
x_3' &= \frac{1}{N_P}\frac{1}{2}x_2x_3 && (4.58)\\
x_4' &= \frac{1}{N_P}(x_4^2 + x_2x_4 + \frac{1}{2}x_2x_3) && (4.59)\\
N_P &= (x_1 + x_2)^2 + x_4^2 + 2x_2(x_3 + x_4). && (4.60)
\end{aligned}$$

Looking for fixed points $x_i' = x_i =: x_i^*$, incorporating conditions $x_i \in \mathbb{R}[0, 1]$,

and $x_3 = 0$, leads to the following six fixed points $\mathbf{x}^* = [x_1^*, x_2^*, x_3^*, x_4^*]$:

$$\mathbf{x}_1^* = [1,0,0,0] \tag{4.61}$$
$$\mathbf{x}_2^* = [0,1,0,0] \tag{4.62}$$
$$\mathbf{x}_3^* = [0,0,0,1] \tag{4.63}$$
$$\mathbf{x}_4^* = [x_1^*, x_2^*, 0, 0] \tag{4.64}$$
$$\mathbf{x}_5^* = [0, x_2^*, 0, x_4^*] \tag{4.65}$$
$$\mathbf{x}_6^* = [\tfrac{1}{2}, 0, 0, \tfrac{1}{2}]. \tag{4.66}$$

The fixed points $[1,0,0,0]$, $[0,1,0,0]$, and $[0,0,0,1]$ are special cases of $\mathbf{x}_4^*$ and $\mathbf{x}_5^*$. These latter two fixed points describe two sets of fixed points, where the non-zero components can take arbitrary positive values as long as their sum is 1. In the fixed point $\mathbf{x}_6^*$, no mating between $AM$- and $am$-gametes takes place, they are reproductively isolated.

The process defined by equations (4.56)–(4.60) always converges to one of the two fixed points (4.64) and (4.65). We can construct the Lyapunov-function $V(x_1, x_2, x_3, x_4) = 1 - N_P$ for these fixed points. $V(x_1^*, x_2^*, x_3^*, x_4^*) = 0$ if and only if $[x_1^*, x_2^*, x_3^*, x_4^*]$ has the form of (4.64) or (4.65). Since $N_P \leq 1$, $V(x_1, x_2, x_3, x_4) \geq 0$. Naming the population composition at time $t$ $\mathbf{x}^t = [x_1, x_2, 0, x_4]$ and the normalization factor $N_P^t$, then

$$V(\mathbf{x}^{t+1}) - V(\mathbf{x}^t) = N_P^t - N_P^{t+1} \tag{4.67}$$
$$= \frac{2x_1x_4}{(N_P^t)^2}(5x_1x_4 - 4x_1^2x_4^2 - x_1 - x_4) \leq 0. \tag{4.68}$$

The last factor in equation (4.68) is negative except at $x_1 = x_4 = 1/2$ and $x_1 = x_4 = 0$, where it is zero. Therefore, $V(x_1, x_2, x_3, x_4)$ is a strict Lyapunov-function with respect to the fixed points (4.64) and (4.65). These two sets of fixed points are asymptotically stable. Moreover, since the Lyapunov-function is strictly decreasing, there are no periodic orbits and we know that the population will converge to either $\mathbf{x}_4^*$ or $\mathbf{x}_5^*$. The fixed point $\mathbf{x}_6^*$ is not attracting; the only point in its attraction basin is itself.

The attraction basins for all initial populations with $x_3 = 0$ are shown in figure 4.2. The relative frequency of gamete $x_i$ in generation $t$ is denoted by $p_t(x_i)$. Consequently, the initial frequency of $x_i$ is denoted by $p_0(x_i)$. The borderline between the fixed points $\mathbf{x}_4^*$ and $\mathbf{x}_5^*$ is determined by the functional relation $p_0(AM) = (1 - p_0(Am))/2$. Using $p_0(AM) + p_0(Am) + p_0(am) = 1$, we see that also $p_0(am) = (1 - p_0(Am))/2$. Both relations together imply that at that line also $p_0(AM) = p_0(am)$ holds.

The rates of change $dx_i := (x_i' - x_i)/x_i$ of gamete $x_i$ in subsequent generations help us to understand the pattern of the attraction basins:

$$dx_1 = \frac{2x_1x_4 - x_4}{1 - 2x_1x_4} = \frac{x_1 + x_2 - N_P}{N_P} \tag{4.69}$$
$$dx_2 = \frac{2x_1x_4}{1 - 2x_1x_4} = \frac{1 - N_P}{N_P} \tag{4.70}$$
$$dx_4 = \frac{2x_1x_4 - x_1}{1 - 2x_1x_4} = \frac{x_2 + x_4 - N_P}{N_P}. \tag{4.71}$$

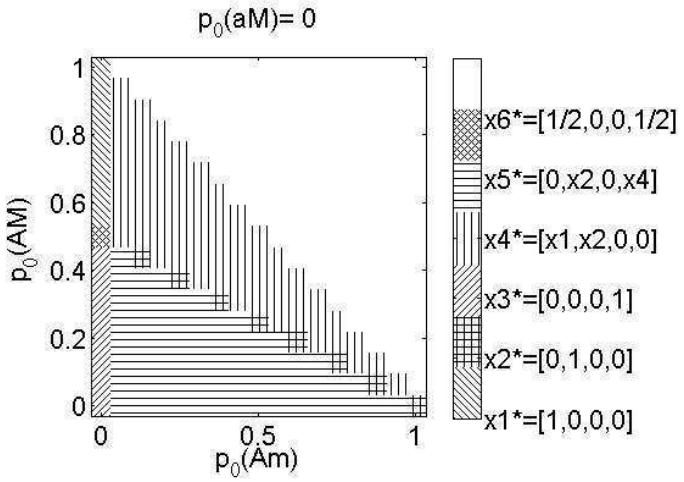



Figure 4.2: The attraction basins of the fixed points in dependence on the initial population composition. The start vector is set to $[x_1, x_2, 0, 1-x_1-x_2]$. The y-axis shows the initial values of gamete $AM$ and the x-axis those of $Am$. The axis are subdivided in steps of $\frac{1}{16}$.

We see that $dx_1, dx_4 < dx_2$ and $dx_2 > 0$ when $x_1, x_4 > 0$. The latter implies that the relative frequency of gamete $Am$ steadily increases as long as $x_1, x_4 > 0$. Furthermore, $dx_1 > 0 \Leftrightarrow x_1 > 1/2$, which means that also $AM$ increases in this case. Consequently for the case $x_1 > 1/2$, $AM$ and $Am$ gain offspring at cost of $am$, the latter thus decreasing in relative frequency. The result is a population in $\mathbf{x}_4^*$, or if $p_0(AM) = p_0(am) = 1/2$, then a population in $\mathbf{x}_6^*$. In the case when $p_0(AM) \leq 1/2$, the population evolution depends on the relation of $p_0(AM)$ to $p_0(am)$. There are three cases:

i) $p_0(AM) = p_0(am) \Rightarrow dx_1 = dx_4$, both genotypes start at the same value, decimate at the same rate and thus, become extinct at the same time, such that gamete $Am$ remains and the fixed point $\mathbf{x}_2^*$ results.

ii) $p_0(AM) > p_0(am) \Rightarrow 0 > dx_1 > dx_4$, that means $p(am)$ decreases at a faster rate and becomes extinct before $p_t(AM)$ reaches zero. At gamete $am$'s extinction, the population stays in $\mathbf{x}_4^*$.

iii) $p_0(AM) < p_0(am) \Rightarrow dx_1 < dx_4$ and $AM$ becomes extinct before $am$ does. The dynamics remains after the extinction of $AM$ in $\mathbf{x}_5^*$.

Thus, the result here is rather simple. When there are mate preferences without accompanying fitness differences, then the type $aM$, i.e., the one that carries a mate preference incompatible with its own type, will become extinct. Generally, the equilibrium population will have either eliminated the mate preference, or will consist only of preferred types. In any case, this analysis was only meant to prepare the setting for the next section

where I will investigate the interplay between mate preferences and fitness differences, that is, I will ask whether it pays to mate with the fittest partners at the expense of forgoing mating opportunities with less fit partners.

### 4.1.3 Fitness values

The results do not essentially change when there are fitness differences between the carriers of alleles $A$ and $a$, in the sense that matings between $A$-carriers are expected to have more offspring than those between $a$-carriers, with the number of offspring of mixed matings in-between. $aM$-types then not only have the disadvantage of finding fewer mating partners, but also carry the burden of lower fitness. Similarly, $Am$-types still have the best access to mating partners, and in addition they now enjoy a higher fitness themselves. Therefore, again $aM$ goes asymptotically extinct, and either $AM$ or $am$ disappears, except possibly when $Am$ was initially absent. In addition, $Am$ and $am$ can no longer co-exist, as $Am$ now is fitter than $am$. Gamete $Am$ will then eventually dominate the population, since matings between the $Am$ and $am$ are not prevented. $AM$ and $Am$ can co-exist as before. And we still have an unstable equilibrium between $AM$ and $am$ which do not interbreed, but now the equilibrium frequency of $am$ needs to be correspondingly higher in order to compensate for the lower fitness. These predictions can be confirmed by straightforward computations of which I give a short sample.

I will set the fitness $f(x_i)$ of gamete $x_i$ to:

$$f(AM) = f(Am) = 1 + f, \quad f(aM) = f(am) = 1. \tag{4.72}$$

The fitness value of a couple formed by gametes $x_i$ and $x_j$, $F(x_i, x_j)$, is set to the average of their fitness values:

$$F(x_i, x_j) \;=\; \frac{f(x_i) + f(x_j)}{2}, \tag{4.73}$$

which is an assumption that is often made in literature.

The not yet normalized relative frequency of gamete $x_i$ in the following generation is now

$$x_i' = \sum_{(x_j, x_k)} F(x_j, x_k) P_{\text{mate}}(x_j, x_k) P_{\text{off}}(x_i | x_j, x_k), \tag{4.74}$$

with $P_{\text{mate}}$ given in table 4.2 and $P_{\text{off}}$ given in equation (2.3). The term $F$ is defined by

$$F = \sum_{(x_i, x_j)} F(x_i, x_j) P_{\text{meet}}(x_i, x_j). \tag{4.75}$$

and the term $F/N_P$ yields the mean offspring number per couple. The normalization of equation (4.74) by $F/N_P$ ensures that the relative genotype frequencies sum up to 1 and that the term $N_P$ cancels out such that the

system of recurrence equations for gametic frequencies in the subsequent generation is

$$x_1' = \frac{1}{F}\left((1+f)x_1^2 + (1+f)x_1x_2 + \frac{2+f}{4}x_2x_3\right) \tag{4.76}$$

$$x_2' = \frac{1}{F}\left((1+f)x_2^2 + (1+f)x_1x_2 + \frac{2+f}{2}x_2x_4 + \frac{2+f}{4}x_2x_3\right) \tag{4.77}$$

$$x_3' = \frac{1}{F}\frac{2+f}{4}x_2x_3 \tag{4.78}$$

$$x_4' = \frac{1}{F}\left(x_4^2 + \frac{2+f}{2}x_2x_4 + \frac{2+f}{4}x_2x_3\right) \tag{4.79}$$

$$F = (1+f)(x_1+x_2)^2 + x_4^2 + (2+f)x_2(x_3+x_4). \tag{4.80}$$

The process defined by the equations (4.76)–(4.80) has the following five fixed points:

$$\mathbf{x}_1^* = [1,0,0,0] \tag{4.81}$$

$$\mathbf{x}_2^* = [0,1,0,0] \tag{4.82}$$

$$\mathbf{x}_3^* = [0,0,0,1] \tag{4.83}$$

$$\mathbf{x}_4^* = [x_1^*, x_2^*, 0, 0] \tag{4.84}$$

$$\mathbf{x}_6^* = [\tfrac{1}{2+f}, 0, 0, \tfrac{1+f}{2+f}]. \tag{4.85}$$

Constructing a strict Lyapunov-function for the fixed points (4.81)–(4.84) with $V(x_1,x_2,x_3,x_4) = (1+f)(x_1+x_2)+x_4-F$ shows that the respective fixed points are attracting and that there are no periodic orbits.

Figure 4.3 shows the basins of attraction for each of the fixed points. The starting populations have been set to $[x_1, x_2, 0, x_4]$ because $x_3$ will lack in the equilibrium anyway and because in this way a 2D-visualization is possible.

The concept of *effective fitness* [89, 80] helps us to explain the remaining of $a$-alleles in the population even though allele $a$ has a lower fitness value than $A$. The effective fitness $f_{\text{eff}}$ measures besides selection influence also influences of evolutionary operators like mutation, recombination, and—in our case—mate choice. It is defined as follows

$$x_i' = \frac{f_{\text{eff}}(x_i)}{\bar{f}}x_i \tag{4.86}$$

with $\bar{f} = \sum_{x_i} f(x_i)x_i$ being the mean individual fitness. When does $f_{\text{eff}}(x_1) = f_{\text{eff}}(x_4)$ hold? Using equations (4.76), (4.79), and (4.80) with $x_2 = x_3 = 0$ in combination with (4.86), we obtain

$$f_{\text{eff}}(x_1) = f_{\text{eff}}(x_4) \quad\Leftrightarrow\quad x_1 = \frac{1}{2+f}. \tag{4.87}$$

For $x_1 = \frac{1}{2+f}$, the population will stay in the fixed point $\mathbf{x}_6^*$ as can also be seen in figure 4.3.

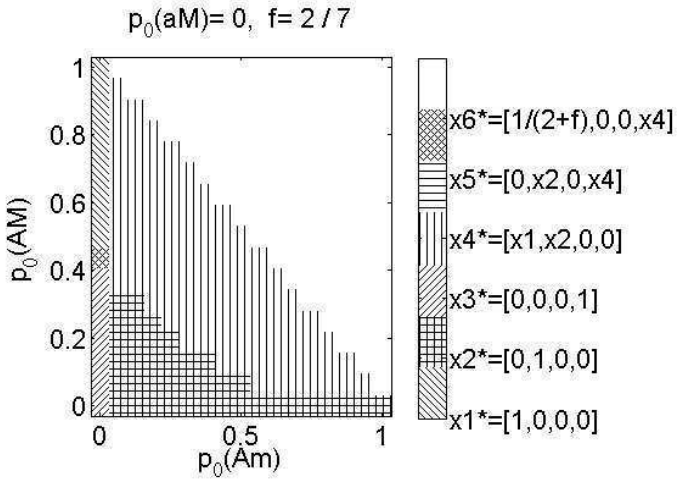



Figure 4.3: The attraction basins of the fixed points in dependence on the initial population composition for $f = \frac{2}{7}$. The start vector is set to $[x_1, x_2, 0, 1 - x_1 - x_2]$. The y-axis shows the initial values of gamete $AM$ and the x-axis those of $Am$.

The borderline between the basins of attraction for $\mathbf{x}_2^*$ and $\mathbf{x}_4^*$ is determined by the functional relation $p_0(AM) = \frac{1}{2}(1 - p_0(Am) - \frac{f}{2+f})$. Using $p_0(AM) + p_0(Am) + p_0(am) = 1$, we can check that at this line also $p_0(AM) = p_0(am) - \frac{f}{2+f}$ holds. Similar to the previous model, where $f$ was zero, one can calculate the change of each gamete in relation to its current frequency:

$$dx_1 = \frac{x_1' - x_1}{x_1} = \frac{x_4}{F}\big((2+f)x_1 - 1\big) \tag{4.88}$$

$$dx_2 = \frac{x_2' - x_2}{x_2} = \frac{x_4}{F}\left((2+f)x_1 + \frac{f}{2}\right) \tag{4.89}$$

$$dx_4 = \frac{x_4' - x_4}{x_4} = \frac{x_4}{F}\left((2+f)x_1x_4 - (1+f)x_1 - \frac{f}{2}x_2\right). \tag{4.90}$$

It holds that $dx_2 > 0$ and $dx_1 < dx_2$ when $x_4 > 0$. If $p_0(AM) > 1/(2+f)$ and $p_0(Am) > 0$, then both gametic frequencies increase, which leads to a population in the fixed point $\mathbf{x}_4^*$. When $p_0(AM) \leq 1/(2+f)$ and $p_0(Am) > 0$ then the dynamics depends on the relation of $dx_1$ to $dx_4$. Analogously to the previous model, there are three cases relating to the borderline:

i) When $p_0(AM) = p_0(am) - f/(2+f) \Rightarrow dx_1 = dx_4$, the population finds its stable point in $\mathbf{x}_2^*$.

ii) If $p_0(AM) > p_0(am) - f/(2+f) \Rightarrow 0 > dx_1 > dx_4$, then the population settles in $\mathbf{x}_4^*$.

iii) $p_0(AM) < p_0(am) - f/(2+f) \Rightarrow dx_1 < dx_4$. Even though gamete $AM$ vanishes faster than gamete $am$, both become extinct, because $Am$ is fitter than $am$. The dynamics ends in $\mathbf{x}_2^*$.

### 4.1.4 Two niches with opposite selection schemes

I will now introduce a second niche. In the existing environment so far, allele $A$ was favored by selection. This environment corresponds now to niche 1. Allele $a$ is favored by selection in the new niche 2 and the fitness values for the gametes in niche 2 are set to

$$f^{\text{niche 2}}(AM) = f^{\text{niche 2}}(Am) = 1, \quad f^{\text{niche 2}}(aM) = f^{\text{niche 2}}(am) = 1+f. \tag{4.91}$$

I have labeled the fitness advantage $f$ with additional indices to specify that niche 2 is meant. Also, I have set the fitness values in such way that the selection pressure is opposite in both niches.

Both niches are linked by migration, and migration takes place prior to mating. A population fraction of each niche migrates to the other niche. The migration rate $p_m$ corresponds to that fraction. I will add an index onto frequency $x_i$ to clarify which niche is meant. Migration changes the relative frequency of $x_i^{\text{niche 1}}$ and $x_i^{\text{niche 2}}$ to

$$x_i^{\text{niche 1}} \mapsto (1-p_m)x_i^{\text{niche 1}} + p_m x_i^{\text{niche 2}} \tag{4.92}$$

$$x_i^{\text{niche 2}} \mapsto (1-p_m)x_i^{\text{niche 2}} + p_m x_i^{\text{niche 1}}. \tag{4.93}$$

Fitness-based mating in niche 2 implies now that $M$-carriers prefer $a$-carriers. The mating probabilities are constructed analogously to those in niche 1, just that the roles of the alleles $A$ and $a$ are exchanged. Table 4.3 shows the mating probabilities in both niches. The normalization factors $N_P^{\text{niche 2}}$ and $F^{\text{niche 2}}$ for the second niche are analogously to the equations (4.46) and (4.80)

$$N_P^{\text{niche 2}} = P(a)^2 + P(Am)^2 + 2P(am)P(A) \tag{4.94}$$

$$F^{\text{niche 2}} = (1+f)P(a)^2 + P(Am)^2 + (2+f)P(am)P(A), \tag{4.95}$$

where $P(x_i)$ denotes the relative gametic frequency after migration. Please note that I have dropped the niche indices for the relative gamete frequencies such that I have shortened both $P^{\text{niche 1}}$ and $P^{\text{niche 2}}$ to $P$, and I have done so for better readability and because it is clear from the context which niche is meant in each equation.

The system of recurrence equations for the gametic frequencies in niche 1 is given by equations (4.76)–(4.80), but we have to replace $x_i$ as in equation (4.92) to account for migration. The equations for niche 2 are similarly constructed by adjusting $F(x_j, x_k)$ using (4.91), by using $P_{\text{mate}}$ from the second column in table 4.3, and by replacing $x_i^{\text{niche 2}}$ by (4.93).

Without migration, we could have an equilibrium population with $AM$ and $Am$-types in niche 1, and with $aM$ and $am$-types in niche 2. When migration occurs, however, then the $AM$-type in niche 1 will be less successful than $Am$, and analogously the $aM$-type in niche 2 will be less successful than $am$, because their mating success in the other niche is lower. So, there will be a higher back-migration of $m$-carriers than of $M$-carriers. Therefore, the effective reproduction of $Am$ in niche 1 is higher than that of $AM$ and the latter type should become extinct. Analogously,

| couple | niche 1 $P_{\text{mate}} \cdot N_P^{\text{niche 1}}$ | niche 2 $P_{\text{mate}} \cdot N_P^{\text{niche 2}}$ |
|---|---|---|
| $AM, AM$ | $P(AM)^2$ | 0 |
| $AM, Am$ | $2P(AM)P(Am)$ | 0 |
| $AM, aM$ | 0 | 0 |
| $AM, am$ | 0 | $2P(AM)P(am)$ |
| $Am, Am$ | $P(Am)^2$ | $P(Am)^2$ |
| $Am, aM$ | $2P(Am)P(aM)$ | 0 |
| $Am, am$ | $2P(Am)P(am)$ | $2P(Am)P(am)$ |
| $aM, aM$ | 0 | $2P(aM)^2$ |
| $aM, am$ | 0 | $2P(aM)P(am)$ |
| $am, am$ | $P(am)^2$ | $P(am)^2$ |

Table 4.3: The mating probabilities in two niches.

$aM$ should disappear in niche 2. For the remaining types, $Am$ and $am$, we then simply need to determine the selection-migration equilibrium. We can solve the system of recurrence equations by setting $x_1 = x_3 = 0$ and $x_4 = 1 - x_2$ in equations (4.76)–(4.80) to obtain the selection-migration equilibrium of $Am$ and $am$. The result is:

$$x_2^{*\text{niche 1}} = \begin{cases} \frac{1}{2}, & \text{if } f = 0 \\ \frac{1}{2} + \frac{2p_m(2+f) - \sqrt{f^2+16p_m^2(1+f)}}{2f(4p_m^2-1)}, & \text{if } f \neq 0, p_m \neq \frac{1}{2} \\ \frac{1}{2} + \frac{f}{4(2+f)}, & \text{if } p_m = \frac{1}{2}. \end{cases} \tag{4.96}$$

In the selection-migration equilibrium of $Am$ and $am$, it holds that

$$x_2^{\text{niche 1}} = x_4^{\text{niche 2}}, \quad x_4^{\text{niche 1}} = x_2^{\text{niche 2}}, \tag{4.97}$$

because both gametes mate randomly and the niche conditions are symmetrical. Thus, we automatically know the values of $x_2^*$ and $x_4^*$ in niche 2. The dependence of $x_2^{*\text{niche 1}}$ on $p_m$ and $f$ is plotted in figure 4.4

Figure 4.5 shows the attraction basins for niche 1 and for starting populations that have the form $[x_1, x_2, 0, x_4]$ in niche 1 and $[0, x_4, x_1, x_2]$ in niche 2. I have chosen these starting populations, because they are especially good for visualizing. Other starting populations have the same dynamics and also settle mostly in the fixed point $\mathbf{x}_5^*$. Figure 4.5 shows that this fixed point prevails. The values of $\mathbf{x}_5^*$ are $[0, x_2^*, 0, x_4^*]$ in niche 1 and $[0, x_4^*, 0, x_2^*]$ in niche 2, where $x_2^*$ has been calculated in equation (4.96). If a starting population has no $m$-carriers, i.e., the population has the form $[x_1, 0, x_3, 0]$ in niche 1 and $[x_1, 0, x_3, 0]$ in niche 2, then the population will settle in the fixed point $\mathbf{x}_1^*$. Its values are $[1, 0, 0, 0]$ in niche 1 and $[0, 0, 1, 0]$ in niche 2. The fitness advantage $f$ and the migration rate $p_m$ have been set to $f = 0$ and $p_m = 0.1$ in figure 4.5. Neither higher $f$-values, nor higher

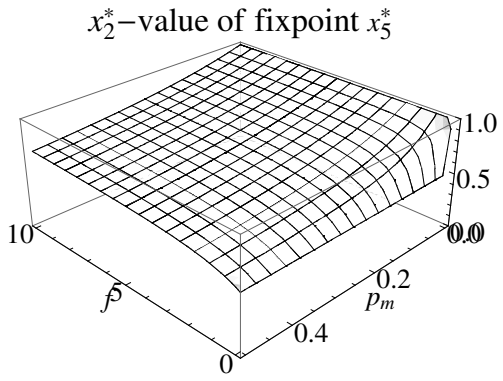



Figure 4.4: The value of $x_2^*$ of the fixed point $\mathbf{x}_5^*$, see equation (4.96), in dependence on $f$ and $p_m$.

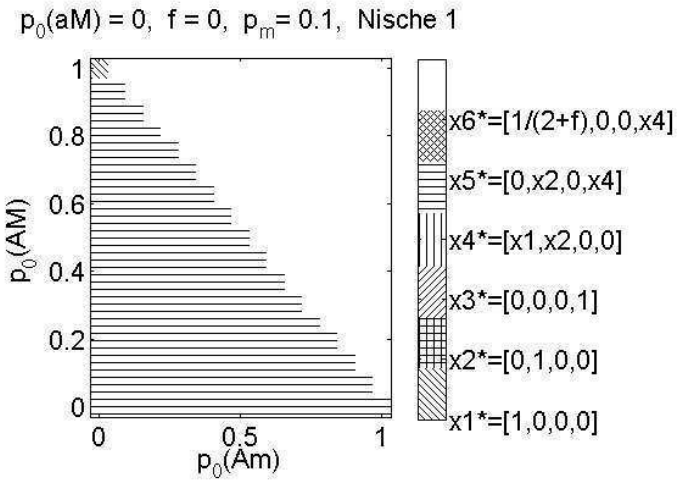



Figure 4.5: The attraction basins of the fixed points in dependence on the initial population composition. The starting populations are set to $[x_1, x_2, 0, 1 - x_1 - x_2]$ in niche 1 and $[0, 1 - x_1 - x_2, x_1, x_2]$ in niche 2. The y-axis shows the initial values of gamete $AM$ and the x-axis those of $Am$. The fitness difference is set to $f = 0$ and migration to $p_m = 0.1$.

$p_m$-values change the dynamics; they lead to exactly the same basins of attraction.

The absence of allele $M$ in the predominant equilibrium is not a desired outcome and I will slightly change the mating condition in the next section, such that more interesting dynamics result.

### 4.1.5 One mating only

So far, the situation was polygamous, or more precisely, matings were not costly, and then, obviously, the best strategy is to mate as often as possible, regardless of the quality of the mates. But the situation changes when we impose that each individual can only mate once. Limiting the number of matings will diminish the disadvantage of $M$. The key point will be that preferential maters will then mate the fittest and random maters will mate an arbitrary partner. In that way, the disadvantage of having less mating chances for preference-maters will be lessened by their fitter partners while random maters will just have average fit partners. This is equivalent to mate preferences without costs, because choosy maters will mate even when the preferred partner is rare.

I will calculate the mean offspring numbers under the condition that each individual can choose exactly one mating partner. But, individuals can mate more than once by being chosen. It is thus possible that an individual takes part in a mating once as the *choosing* mate and several times in other matings as the *chosen* mate. Yet, not *every* mate is accepted. A chosen candidate, when having a mate preference expressed at the second locus, checks whether the choosing candidate satisfies its criterion at the first locus. If so, the choosing candidate is accepted, if not, rejected. So as before, matings are based on mutual agreement[4]. In section 4.1.6, I will give up that requirement.

Let us adjust the formulas to the new situation. As before, $N_P$ is the normalization factor of mating probabilities defined by equation (4.9). But now, $P(x_i)P_{x_i}^{\text{choice}}(x_j)$, which is the product of the relative frequency of $x_i$ and the probability that $x_i$ chooses $x_j$, replaces $P_{\text{meet}}$. The probability that $x_j$ accepts when chosen by $x_i$, namely $P_{x_j}^{\text{accept}}(x_i)$, is as in table 4.4. The mating probability is now set to

$$P_{\text{mate}}(x_i,x_j) = \begin{cases} \frac{1}{N_P}P(x_i)P_{x_i}^{\text{choice}}(x_j)P_{x_j}^{\text{accept}}(x_i), & \text{if } i=j \\ \frac{1}{N_P}\Big(P(x_i)P_{x_i}^{\text{choice}}(x_j)P_{x_j}^{\text{accept}}(x_i)+\dots \\ \quad +P(x_j)P_{x_j}^{\text{choice}}(x_i)P_{x_i}^{\text{accept}}(x_j)\Big), & \text{otherwise.} \end{cases} \tag{4.98}$$

Table 4.5 shows the choosing probabilities and table 4.6 shows the mating probabilities for this setting. The system of recurrence equations for the

[4]The mating with mutual agreement is also known as two-sided matching in economics literature, see review [5].

gametic frequencies is normalized by $F/N_P$ with $F$ defined as

$$F = \sum_{(x_i,x_j)} F(x_i,x_j)P_{\text{mate}}(x_i,x_j)N_P. \tag{4.99}$$

The normalization factors $N_P$ in niche 1 and 2 are:

$$N_P^{\text{niche 1}} = \begin{cases} \frac{P(AM)^2}{P(A)} + P(m)^2 + P(Am)P(M)\left(1+\frac{1}{P(A)}\right) & \text{if } P(A)>0 \\ P(am)^2 & \text{if } P(A)=0 \end{cases} \tag{4.100}$$

$$N_P^{\text{niche 2}} = \begin{cases} \frac{P(aM)^2}{P(a)} + P(m)^2 + P(am)P(M)\left(1+\frac{1}{P(a)}\right) & \text{if } P(a)>0 \\ P(Am)^2 & \text{if } P(a)=0. \end{cases} \tag{4.101}$$

Please note that again I have shortened $P^{\text{niche 1}}$ to $P$ in equation (4.100) and likewise, I have shortened $P^{\text{niche 2}}$ to $P$ in equation (4.101). The respective niche indices are clear from the context and can be dropped for better readability. I will use this convention also in the sequel when the contexts are equally clear. The terms $F$ in niche 1 and 2 calculate, with $x_A = x_1+x_2$, $x_a = x_3+x_4$, and $x_i$ being the relative gamete frequency after migration to:

$$F^{\text{niche 2}} = \begin{cases} (1+f)\left[\frac{x_1^2}{x_A} + x_1x_2\left(1+\frac{1}{x_A}\right) + x_2^2\right] + \dots \\ \quad +\frac{2+f}{2}x_2\left[x_3\left(1+\frac{1}{x_A}\right) + 2x_4\right] + x_4^2 & \text{if } P(A)>0 \\ x_4^2 & \text{if } P(A)=0, \end{cases} \tag{4.102}$$

$$F^{\text{niche 2}} = \begin{cases} (1+f)\left[\frac{x_3^2}{x_a} + x_3x_4\left(1+\frac{1}{x_a}\right) + x_4^2\right] + \dots \\ \quad +\frac{2+f}{2}x_4\left[x_1\left(1+\frac{1}{x_a}\right) + 2x_2\right] + x_2^2 & \text{if } P(a)>0 \\ x_2^2 & \text{if } P(a)=0. \end{cases} \tag{4.103}$$

We see in the equations (4.100)–(4.101) and also in (4.102)–(4.103) that the population dynamics has a discontinuity. For instance, when an $A$-carrier arrives in niche 1 where only $a$-carriers were formerly present, then $N_P^{\text{niche 1}}$ has a point of discontinuity. This is caused by choosy $M$-females that would not mate if $A$-males were absent, but do mate when an $A$-carrier arrives. This is a consequence from no-cost choosiness.

With these terms, the system of recurrence equations for gametic frequencies from one generation to the next is constructed for both niches according to

$$x_i' = \frac{N_P}{F}\sum_{(x_j,x_k)} F(x_j,x_k)P_{\text{mate}}(x_j,x_k)P_{\text{off}}(x_i|x_j,x_k). \tag{4.104}$$

Let us first look at the case without migration and exemplarily discuss the situation in niche 1. Even with limiting the number of matings to one,

| | niche 1 | | | | niche 2 | | | |
|---|---|---|---|---|---|---|---|---|
| | $x_j$ | | | | $x_j$ | | | |
| $x_i$ | $AM$ | $Am$ | $aM$ | $am$ | $AM$ | $Am$ | $aM$ | $am$ |
| $AM$ | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 |
| $Am$ | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 |
| $aM$ | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 |
| $am$ | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 |

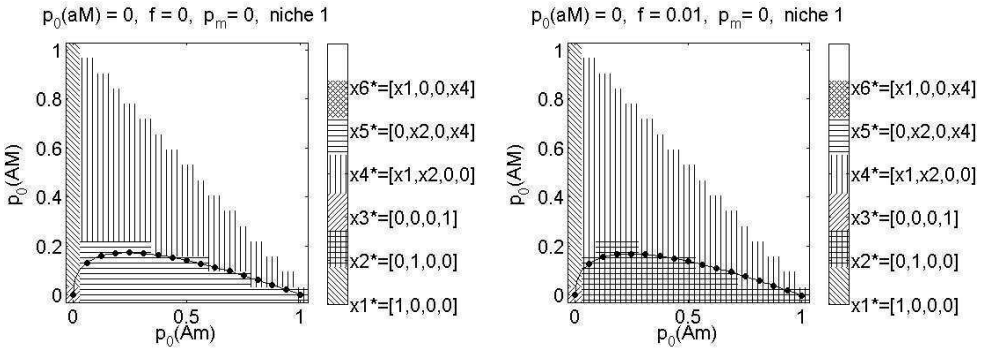
Table 4.4: The acceptance probabilities $P_{x_j}^{\text{accept}}(x_i)$ for two niches

| | niche 1 | | | | niche 2 | | | |
|---|---|---|---|---|---|---|---|---|
| | $x_j$ | | | | $x_j$ | | | |
| $x_i$ | $AM$ | $Am$ | $aM$ | $am$ | $AM$ | $Am$ | $aM$ | $am$ |
| $AM$ | $\frac{P(AM)}{P(A)}$ | $\frac{P(Am)}{P(A)}$ | 0 | 0 | 0 | 0 | $\frac{P(aM)}{P(a)}$ | $\frac{P(am)}{P(a)}$ |
| $Am$ | $P(AM)$ | $P(Am)$ | $P(aM)$ | $P(am)$ | $P(AM)$ | $P(Am)$ | $P(aM)$ | $P(am)$ |
| $aM$ | $\frac{P(AM)}{P(A)}$ | $\frac{P(Am)}{P(A)}$ | 0 | 0 | 0 | 0 | $\frac{P(aM)}{P(a)}$ | $\frac{P(am)}{P(a)}$ |
| $am$ | $P(AM)$ | $P(Am)$ | $P(aM)$ | $P(am)$ | $P(AM)$ | $P(Am)$ | $P(aM)$ | $P(am)$ |

Table 4.5: The choosing probabilities $P_{x_i}^{\text{choice}}(x_j)$ for two niches

| couple | niche 1<br>$P_{\text{mate}} \cdot N_P^{\text{niche 1}}$ | niche 2<br>$P_{\text{mate}} \cdot N_P^{\text{niche 2}}$ |
|---|---|---|
| $AM, AM$ | $\frac{P(AM)^2}{P(A)}$ | 0 |
| $AM, Am$ | $P(AM)P(Am)\left(1+\frac{1}{P(A)}\right)$ | 0 |
| $AM, aM$ | 0 | 0 |
| $AM, am$ | 0 | $P(AM)P(am)\left(1+\frac{1}{P(a)}\right)$ |
| $Am, Am$ | $P(Am)^2$ | $P(Am)^2$ |
| $Am, aM$ | $P(Am)P(aM)\left(1+\frac{1}{P(A)}\right)$ | 0 |
| $Am, am$ | $2P(Am)P(am)$ | $2P(Am)P(am)$ |
| $aM, aM$ | 0 | $\frac{P(aM)^2}{P(a)}$ |
| $aM, am$ | 0 | $P(aM)P(am)\left(1+\frac{1}{P(a)}\right)$ |
| $am, am$ | $P(am)^2$ | $P(am)^2$ |

Table 4.6: The mating probabilities in both niches when only one mating is allowed.




(a) $f = 0$ (b) $f = 0.01$

Figure 4.6: One mating only, no migration. The starting populations are set to $[x_1, x_2, 0, 1 - x_1 - x_2]$ in niche 1. The y-axis shows the initial values of gamete $AM$ and the x-axis those of $Am$. The fitness difference is set to $f = 0$ (left plot) and $f = 0.01$ (right plot), and the migration rate is set to $p_m = 0$.

the gamete $aM$ will find the fewest mating partners, and so it will lack in all equilibria. If $p_m = 0$ and $f = 0$, we will either get the co-existence of $AM$ and $Am$ (fixed point $\mathbf{x}_4^*$), or the co-existence of $Am$ and $am$ (fixed point $\mathbf{x}_5^*$). The gametes of both equilibria have the same effective reproduction rates, because they mate randomly among each other. And if $p_m = 0$ and $f > 0$, we will either get a co-existence of $AM$ and $Am$ (again fixed point $\mathbf{x}_4^*$), or the fixation of $Am$ (fixed point $\mathbf{x}_2^*$). I will plot attraction basins for initial populations that already lack gamete $aM$, because $aM$ will not be present in the steady state and because the basins can then be visualized in 2D. Figure 4.6 shows the attraction basins when each element mates only once and when $P_0(aM) = 0$. The fixed points $\mathbf{x}_4^*$ and $\mathbf{x}_5^*$ (when $f = 0$), and the fixed points $\mathbf{x}_4^*$ and $\mathbf{x}_2^*$ (when $f > 0$) dominate. The borderline between them is found when calculating $dx_i$. They are for $p_m = 0$ and $P_0(aM) = 0$

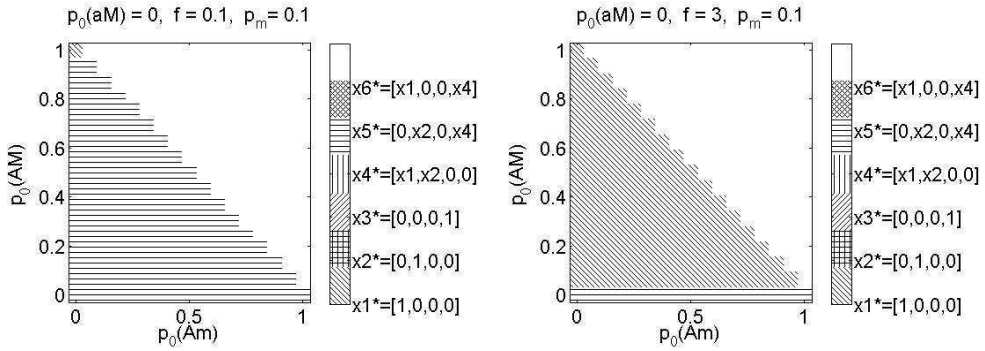
$$dx_1 = \frac{1}{F}\left\{(1+f)\left[1+\frac{1}{2}x_2\left(1-\frac{1}{x_A}\right)\right]-F\right\} \tag{4.105}$$

$$dx_2 = \frac{1}{F}\left\{(1+f)\left[x_2+\frac{1}{2}x_1\left(1+\frac{1}{x_A}\right)\right]+\left(1+\frac{f}{2}\right)x_4-F\right\} \tag{4.106}$$

$$dx_4 = \frac{1}{F}\left(x_2+x_4+\frac{f}{2}x_2-F\right). \tag{4.107}$$

Asking when $dx_1 = dx_4$ gives $2x_1^2+3x_1x_2+x_2^2-x_2 = -f(2x_1+x_2)$. Requiring that $x_1, x_2 \in \mathbb{R}[0,1]$ gives the borderline

$$x_1 = -\frac{1}{4}(2f+3x_2)+\frac{1}{4}\sqrt{4f(x_2+f)+x_2(8+x_2)}. \tag{4.108}$$




(a) $f = 0.1$ (b) $f = 3$

Figure 4.7: One mating only, with migration. The starting populations are set to $[x_1, x_2, 0, 1-x_1-x_2]$ in niche 1 and $[0, 1-x_1-x_2, x_1, x_2]$ in niche 2. The y-axis shows the initial values of gamete $AM$ and the x-axis those of $Am$. The fitness difference is set to $f = 0.1$ (left plot) and $f = 3$ (right plot), and the migration rate is set to $p_m = 0.1$.

This line is plotted above the attraction basins in figure 4.6. As before, we can distinguish three cases as whether $P_0(AM)$ lies above, below, or on the borderline: If the initial relative frequency of $AM$ lies above the borderline, then the dynamics ends in $\mathbf{x}_4^*$. If $P_0(AM)$ lies below that line, then the population will settle in $\mathbf{x}_5^*$ (if $f = 0$) or $\mathbf{x}_2^*$ (if $f > 0$). If $P_0(AM)$ lies on the borderline, then the dynamics will come to its steady state in $\mathbf{x}_2^*$.

Let us now look at the case with migration. We will either get the fixation of $m$ or the fixation of $M$, and this depends on the fitness advantage $f$. If $f$ is sufficiently high, ecological selection will compensate the lower mating success of $M$-types across both niches by higher reproduction rates in one niche. If $M$ goes to fixation, then the fixed point values will be $[1, 0, 0, 0]$ in niche 1 and $[0, 0, 1, 0]$ in niche 2. This is because $AM$ and $aM$ do not mate with each other, since either the one or the other does not match the preferred type of the choosing partner. And so, $AM$ will go to fixation in niche 1, and $aM$ will go to fixation in niche 2. If allele $m$ goes to fixation, the population will settle in the selection-migration equilibrium given by equation (4.96). When the starting populations of both niches consists of only $A$-carriers, i.e., the initial populations have the form $[x_1^{\text{niche 1}}, x_2^{\text{niche 1}}, 0, 0]$ in niche 1 and $[x_1^{\text{niche 2}}, x_2^{\text{niche 2}}, 0, 0]$ in niche 2, then $AM$ will become extinct because it has the same reproduction rate as the $Am$-type in one niche but it forgoes mating chances in the other niche. Consequently, $Am$ will remain in the population. The same is true when the starting populations of both niches consists of only $a$-carriers. In that case, $am$ will remain in the population and $aM$ will disappear.

I will visualize the attraction basins for initial populations that have

the form $[x_1, x_2, 0, x_4]$ in niche 1 and $[0, x_4, x_1, x_2]$ in niche 2, because these populations allow a 2D-visualization and settle in one of the more interesting fixed points. The 1-gamete fixed points, where either $Am$ or $am$ remain as the only gametes, are therefore not visualized, because they are not approached by these initial populations. Figure 4.7 exemplarily shows that a low $f$ leads to the fixation of $m$, and a high $f$ leads to the fixation of $M$ for these initial populations.

So, ecological selection can compensate the lower reproduction rates of $M$-carriers. We will see in the following section that the fixation of $M$-types becomes unconditional if we allow different mating strategies of females and males. In doing so we can account for the different reproductive investments of different sexes.

### 4.1.6 Genders and no mutual agreement

The situation becomes more interesting and biologically more realistic if we introduce genders (female and male) with different mating strategies. Let us assume that males mate indiscriminately and try to achieve as many matings as possible, whereas females can only mate once. I will model this situation as follows: It will be no longer possible for the chosen mate (the male) to reject a mating. Instead, the chosen mate will accept any candidate (any female) even when she does not meet its preference, i.e., I will now omit the mutual mate choice.

Biologically, this corresponds to choosy females and non-choosy males.[5] This mating procedure is applicable to species where females initiate mating. That is the case, for instance, when the mating decision is made after the courtship of several males or after male competition. I need to assume that the mating decision does not involve an harassment of the female. Each female chooses only once, whereas males remain in the mating pool and can mate more than once if they are chosen more than once by different females. Thus, the population is monoandrous and polygynous.

The distribution of the gametes is the same in both genders, meaning that the frequency of gamete $x_i$ is the same in both sexes.[6] Mate preferences are expressed in females only. The acceptance probabilities $P_{x_j}^{\text{accept}}(x_i)$ are set to 1 for all $x_i$ and $x_j$. Thus, the mating probability $P_{\text{mate}}$ depends on the gametic frequencies and on $P^{\text{choice}}$ only; the latter is given in table 4.5. We can then calculate the explicit mating probabilities according to equation (4.98) and they are listed in table 4.7. The normalization factors are then

$$N_P^{\text{niche 1}} = \begin{cases} 1, & \text{if } P(A) > 0 \\ P(am), & \text{if } P(A) = 0, \end{cases} \qquad N_P^{\text{niche 2}} = \begin{cases} 1, & \text{if } P(a) > 0 \\ P(Am), & \text{if } P(a) = 0, \end{cases}$$

[5] On the basis of Batesman's analysis [4] about gender-specific investments into germ cells, where females have higher costs producing germ cells than males, it is often assumed that females are the choosy gender. However, here this is an arbitrary assumption which does not restrict the generality of the model as long as only one sex is choosy.

[6] This is equivalent to assuming loci to be autosomal, which is usually done in population-genetic models, e.g., in [50, 81].

| couple | niche 1<br>$P_{\text{mate}} \cdot N_P^{\text{niche 1}}$ | niche 2<br>$P_{\text{mate}} \cdot N_P^{\text{niche 2}}$ |
|---|---|---|
| $AM, AM$ | $\frac{P(AM)^2}{P(A)}$ | $0$ |
| $AM, Am$ | $P(AM)P(Am)\left(1+\frac{1}{P(A)}\right)$ | $P(AM)P(Am)$ |
| $AM, aM$ | $\frac{P(AM)P(aM)}{P(A)}$ | $\frac{P(AM)P(aM)}{P(a)}$ |
| $AM, am$ | $P(AM)P(am)$ | $P(AM)P(am)\left(1+\frac{1}{P(a)}\right)$ |
| $Am, Am$ | $P(Am)^2$ | $P(Am)^2$ |
| $Am, aM$ | $P(Am)P(aM)\left(1+\frac{1}{P(A)}\right)$ | $P(Am)P(aM)$ |
| $Am, am$ | $2P(Am)P(am)$ | $2P(Am)P(am)$ |
| $aM, aM$ | $0$ | $\frac{P(aM)^2}{P(a)}$ |
| $aM, am$ | $P(aM)P(am)$ | $P(aM)P(am)\left(1+\frac{1}{P(a)}\right)$ |
| $am, am$ | $P(am)^2$ | $P(am)^2$ |

Table 4.7: The mating probabilities in both niches when females are choosy and males are non-choosy.

(4.109)

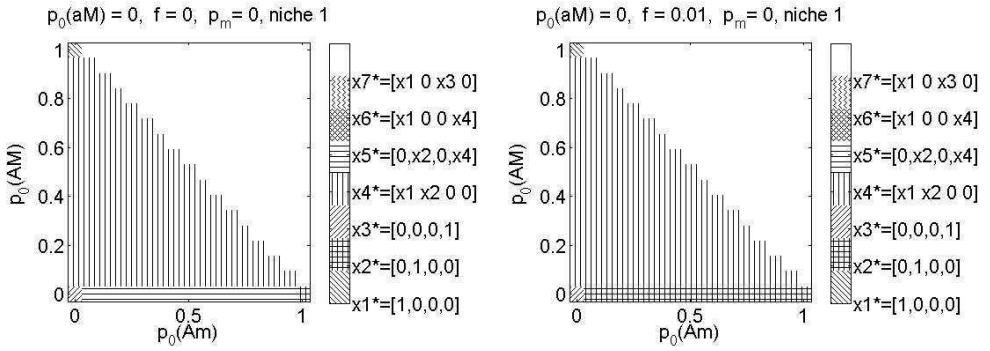
and

$$F^{\text{niche 1}} = \begin{cases} 1+\frac{f}{2}\Big(2P(A)+P(a)P(M)\Big), & \text{if } P(A)>0 \\ P(am), & \text{if } P(A)=0. \end{cases} \tag{4.110}$$

$$F^{\text{niche 2}} = \begin{cases} 1+\frac{f}{2}\Big(2P(a)+P(A)P(M)\Big), & \text{if } P(a)>0 \\ P(Am), & \text{if } P(a)=0. \end{cases} \tag{4.111}$$

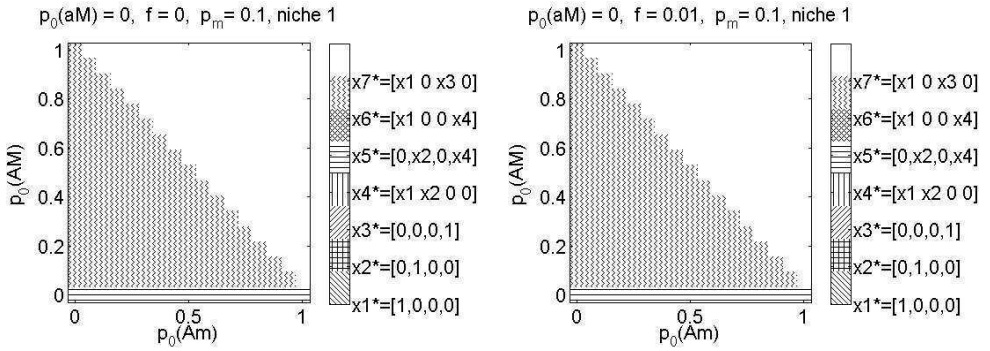
The system of recurrence equations for the gametic evolution can then be constructed as before from equation (4.104).

Let us first look at the case without migration and exemplarily discuss the situation in niche 1. I will compare the offspring distribution of females and the attractiveness of males. An $AM$-female chooses among $AM$-males and $Am$-males, so at least half of its offspring and at most all offspring will be of the same genotype as her. An $aM$-female also chooses among $AM$ and $Am$-males, but at least a quarter and at most a half of its offspring will be of the same type as their mother. Females of the remaining types, namely, $Am$ and $am$, choose among all genotypes, thus, the fraction of their same-type offspring ranges between a quarter and one. In this way, an $AM$-female will have the most "true" offspring and an $aM$-female the least. $Am$ and $am$-females have true-offspring reproduction rates that lie between those of $AM$ and $aM$. Now, I will compare the attractiveness of males. While $AM$-males and $Am$-males are chosen by females of all types, $aM$ and $am$-males are chosen by $m$-females only. All together, the true-offspring production rates of females and the attractiveness of males, allow us to order the effective reproduction rates. The $AM$-type will proliferate




(a) $f = 0$ (b) $f = 0.01$

Figure 4.8: Genders, no migration. The starting populations are set to $[x_1, x_2, 0, 1 - x_1 - x_2]$ in niche 1. The y-axis shows the initial values of gamete $AM$ and the x-axis those of $Am$. The fitness difference is set to $f = 0$ (left plot) and $f = 0.01$ (right plot), and the migration rate is set to $p_m = 0$.




(a) $f = 0$ (b) $f = 0.01$

Figure 4.9: Genders, with migration. The starting populations are set to $[x_1, x_2, 0, 1 - x_1 - x_2]$ in niche 1 and $[0, 1 - x_1 - x_2, x_1, x_2]$ in niche 2. The y-axis shows the initial values of gamete $AM$ and the x-axis those of $Am$. The fitness difference is set to $f = 0$ (left plot) and $f = 0.01$ (right plot), and the migration rate is set to $p_m = 0.1$.

the most because it has both the most true-offspring as a female and the most mating partners as a male. The $Am$-type has the second highest effective reproduction rate because it breeds true less often than $AM$ and is as attractive as $AM$. The $am$-type breeds as often true as $Am$ but is less attractive than $Am$. Therefore, $am$ has the third highest effective reproduction rate. The $aM$-type has the lowest reproduction rates because it breeds true less often than any other type and it is as less attractive as $am$. Once $a$-carriers will have become extinct, the population will come to its steady state because the remaining types, namely, $AM$ and $Am$, have in the absence of $a$ the same true-breeding rates and the same attractiveness onto each other. So, the equilibrium will be a co-existence of $AM$ and $Am$-types. Figure 4.8 shows the attraction basins when starting populations already lack $aM$-types. These starting populations have been chosen because they allow a 2D-visualization. Furthermore, they are representative in the sense that also different starting populations settle mostly—except when more than one type is already initially absent—in an equilibrium of only $A$-types. We see in figure 4.8 that the fixed point $\mathbf{x}_4^{*\text{niche1}} = [x_1^*, x_2^*, 0, 0]$ prevails.

Before turning to the case with migration, straightforward calculations will show that the relative frequency of allele $M$ increases from one generation to the next:

$$p'(M) = p'(AM) + p'(aM) \tag{4.112}$$

$$= -\frac{(x_1 + x_3)[x_1^2 + x_2(x_3 - 2) + x_1(x_2 + x_3 - 3)]}{2(x_1 + x_2)}. \tag{4.113}$$

If we ask when $p'(M) \geq p(M)$, we get the condition

$$x_2 \leq \frac{x_1(x_2 + x_4)}{x_1 + x_3}. \tag{4.114}$$

And because it holds that

$$\frac{x_1(x_2 + x_4)}{x_1 + x_3} \leq x_2 + x_4, \tag{4.115}$$

we know that $p'(M) \geq p(M)$ is indeed fulfilled. The relative frequency of allele $M$ increases from one generation to the next, and this frequency-increase continues as long as $x_3 + x_4 > 0$, i.e., as long as allele $A$ did not yet go to fixation.

The preceding discussion helps us to also understand the dynamics when there is migration. The relative frequency of $M$ increases in both niches. Additionally we know that $AM$ has the highest reproductive success in niche 1, and $aM$ has the highest reproductive success in niche 2. The equilibrium therefore must consist of $AM$ and $aM$ only. I will plot the attraction basins exemplarily for starting populations that have the form $[x_1, x_2, 0, x_4]$ in niche 1 and have the form $[0, x_4, x_1, x_2]$ in niche 2. Again I have chosen these starting populations because they especially suit a 2D-visualization. But also any other initial condition leads to the same

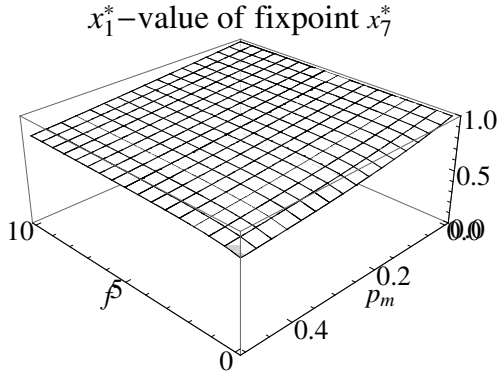



Figure 4.10: No mutual agreement, genders. The value of $x_1^*$ of the fixed point $\mathbf{x}_7^*$ in dependence on $f$ and $p_m$, see equation (4.116).

dynamics. Figure 4.9 shows the attraction basins when there is migration between the niches and we see that the population settles in the new fixed point $\mathbf{x}_7^{*\text{niche1}} = [x_1^*, 0, x_3^*, 0]$. The value of $x_1^{*\text{niche 1}}$ of the fixed point $\mathbf{x}_7^{*\text{niche 1}}$ can be calculated:

$$x_1^{*\text{niche 1}} = \begin{cases} \frac{1+p_m}{1+2p_m}, & \text{if } p_m \in (0, \frac{1}{2}), f = 0 \\ \frac{2-f+4p_m(1+2f)}{4f(2p_m-1)} - \dots \\ -\frac{\sqrt{(-2+f-4p_m(1-2f))^2-8f(2+f+p_m(2+3f))(2p_m-1)}}{4f(2p_m-1)}, & \text{if } p_m \in (0, \frac{1}{2}), f \neq 0 \\ \frac{6+5f}{8+6f}, & \text{if } p_m = \frac{1}{2}, f \neq 0 \end{cases} \tag{4.116}$$

We also know from equation (4.116) which value $x_3^{*\text{niche 2}}$ has, because the niche conditions are symmetric and the roles of $AM$ and $aM$ are exchanged in niche 2. Therefore, it holds that $x_1^{*\text{niche 1}} = x_3^{*\text{niche 2}}$. Figure 4.10 shows how $x_1^{*\text{niche 1}}$ depends on the fitness advantage $f$ and migration rate $p_m$. The value of $x_1^*$ decreases with $p_m$ and increases slightly with $f$.

In conclusion, fitness-based mating spreads in the population when migration is present. If migration is absent, the ecologically favored allele goes to fixation such that there is no difference in the mating behavior of $M$- and $m$-females. The key assumptions were that the mate preference is expressed by females only and that the population is monoandrous and polygynous.

### 4.1.7 Tendency for fitness-based mating

I will now introduce a tendency for fitness-based mating. So far, $M$-females have preferred exclusively the fittest mates. Now, a parameter $m_k$ is introduced that gives the probability that a female will mate fitness-based. The index $k$ counts the number of $M$-alleles in the gamete, so $k$

| female $x_i$ | male $x_j$ | niche 1 $P_{x_i}^{\text{choice}}(x_j)$ | niche 2 |
|---|---|---|---|
| $AM$ | $AM$ | $m_1\frac{P(AM)}{P(A)}+(1-m_1)P(AM)$ | $(1-m_1)P(AM)$ |
| | $Am$ | $m_1\frac{P(Am)}{P(A)}+(1-m_1)P(Am)$ | $(1-m_1)P(Am)$ |
| | $aM$ | $(1-m_1)P(aM)$ | $m_1\frac{P(aM)}{P(a)}+(1-m_1)P(aM)$ |
| | $am$ | $(1-m_1)P(am)$ | $m_1\frac{P(am)}{P(a)}+(1-m_1)P(am)$ |
| $Am$ | $AM$ | $P(AM)$ | $P(AM)$ |
| | $Am$ | $P(Am)$ | $P(Am)$ |
| | $aM$ | $P(aM)$ | $P(aM)$ |
| | $am$ | $P(am)$ | $P(am)$ |
| $aM$ | $AM$ | $m_1\frac{P(AM)}{P(A)}+(1-m_1)P(AM)$ | $(1-m_1)P(AM)$ |
| | $Am$ | $m_1\frac{P(Am)}{P(A)}+(1-m_1)P(Am)$ | $(1-m_1)P(Am)$ |
| | $aM$ | $(1-m_1)P(aM)$ | $m_1\frac{P(aM)}{P(a)}+(1-m_1)P(aM)$ |
| | $am$ | $(1-m_1)P(am)$ | $m_1\frac{P(am)}{P(a)}+(1-m_1)P(am)$ |
| $am$ | $AM$ | $P(AM)$ | $P(AM)$ |
| | $Am$ | $P(Am)$ | $P(Am)$ |
| | $aM$ | $P(aM)$ | $P(aM)$ |
| | $am$ | $P(am)$ | $P(am)$ |

Table 4.8: The choosing probabilities under tendential fitness-based mating in the absence of veto-rights.

can take the values $k=0$ and $k=1$. Thus, $m$-females mate fitness-based with a probability of $m_0$ and mate randomly with a probability of $1-m_0$. Likewise, $M$-females will do so with a probability of $m_1$ respectively of $1-m_1$. I will assume that $m_0=0$ and that $m_1\geq 0$, such that fitness-based mating competes with random mating. When we set $m_1=1$, we get the model of section 4.1.6, and when we set $m_1=0$, the population mates randomly.

The new choosing probabilities are given in table 4.8. Using these choosing probabilities, we can calculate the mating probabilities by formula (4.98). They are listed in table 4.9. The normalization factors for $P_{\text{mate}}$ and for the system of recurrence equations are

$$N_P^{*\text{niche 1}} = \begin{cases} 1, & \text{if } P(A)>0 \\ 1-m_1P(aM), & \text{if } P(A)=0, \end{cases} \tag{4.117}$$

$$N_P^{*\text{niche 2}} = \begin{cases} 1, & \text{if } P(a)>0 \\ 1-m_1P(AM), & \text{if } P(a)=0, \end{cases} \tag{4.118}$$

and

$$F^{*\text{niche 1}} = \begin{cases} 1+f\left(P(A)+\frac{m}{2}P(a)P(M)\right), & \text{if } P(A)>0 \\ 1-m_1P(aM), & \text{if } P(A)=0, \end{cases} \tag{4.119}$$

$$F^{*\text{niche 2}} = \begin{cases} 1+f\left(P(a)+\frac{m}{2}P(A)P(M)\right), & \text{if } P(a)>0 \\ 1-m_1P(AM), & \text{if } P(a)=0. \end{cases} \tag{4.120}$$

| couple | niche 1<br>$P_{\text{mate}} \cdot N_P^{\text{niche 1}}$ | niche 2<br>$P_{\text{mate}} \cdot N_P^{\text{niche 2}}$ |
|---|---|---|
| $AM, AM$ | $P(AM)^2\left(1-m_1+\frac{m_1}{P(A)}\right)$ | $P(AM)^2(1-m_1)$ |
| $AM, Am$ | $P(AM)P(Am)\left(2-m_1+\frac{m_1}{P(A)}\right)$ | $P(AM)P(Am)(2-m_1)$ |
| $AM, aM$ | $P(AM)P(aM)\left(2(1-m_1)+\frac{m_1}{P(A)}\right)$ | $P(AM)P(aM)\left(2(1-m_1)+\frac{m_1}{P(a)}\right)$ |
| $AM, am$ | $P(AM)P(am)(2-m_1)$ | $P(AM)P(am)\left(2-m_1+\frac{m_1}{P(a)}\right)$ |
| $Am, Am$ | $P(Am)^2$ | $P(Am)^2$ |
| $Am, aM$ | $P(Am)P(aM)\left(2-m_1+\frac{m_1}{P(A)}\right)$ | $P(Am)P(aM)(2-m_1)$ |
| $Am, am$ | $2P(Am)P(am)$ | $2P(Am)P(am)$ |
| $aM, aM$ | $P(aM)^2(1-m_1)$ | $P(aM)^2\left(1-m_1+\frac{m_1}{P(a)}\right)$ |
| $aM, am$ | $P(aM)P(am)(2-m_1)$ | $P(aM)P(am)\left(2-m_1+\frac{m_1}{P(a)}\right)$ |
| $am, am$ | $P(am)^2$ | $P(am)^2$ |

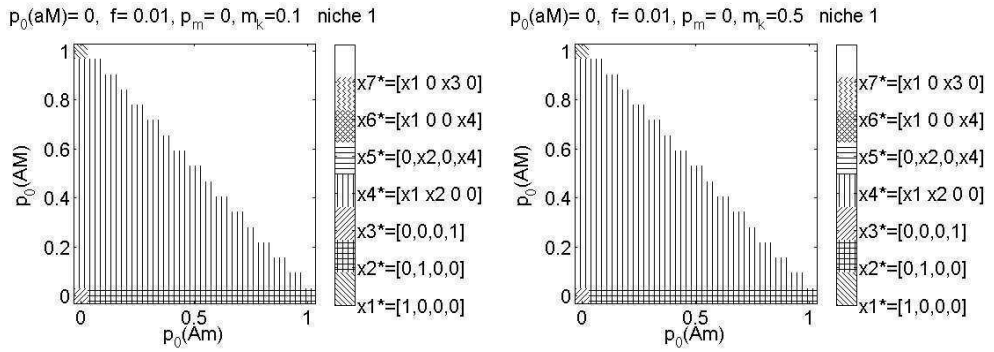
Table 4.9: The mating probabilities in both niches under tendential fitness-based mating.

The population dynamics follows exactly the reasoning outlined in section 4.1.6. Without migration, the fittest 1-locus allele goes to fixation in each niche, and with migration, allele $M$ goes to fixation. Also figures 4.11 (without migration) and figures 4.12 (with migration) show for different $m_k$-values that there is no significant change in the attraction basins compared to the previous model. The values of $\mathbf{x}_5^*$ are exactly as in equation (4.96), since the mating probabilities of $P_{\text{mate}}(Am, Am)$, $P_{\text{mate}}(Am, am)$, and $P_{\text{mate}}(am, am)$ do not depend on $m_1$. Under migration, the equilibrium in which the population settles for most initial conditions is $\mathbf{x}_7^{*\text{niche 1}} = [x_1^*, 0, x_3^*, 0]$ and $x_1^{*\text{niche 1}}$ of $\mathbf{x}_7^{*\text{niche 1}}$ can be calculated:

$$x_1^{*\text{niche 1}} = \begin{cases} \frac{1}{2C}\left(D+\sqrt{D^2-4EC}\right) & \text{if } p_m \in (0,\frac{1}{2}), f \neq 0, m_1 > 0 \\ \text{with} & \\ C{=}2f(1{-}2m_1p_m{-}(1{-}m_1)4p_m^2) & \\ D{=}2(2p_m(m_1{-}2){-}m_1){-}f(m_1{-}2{+}4p_m(1{+}m_1){+}8p_m^2(1{-}m_1)) & \\ E{=}2(p_m(m_1{-}2){-}m_1){-}f(m_1{+}p_m(2{+}m_1){+}2p_m^2(1{-}m_1)) & \\ \frac{1}{2}+\frac{m_1}{2m_1+4p_m(2-m_1)}, & \text{if } p_m \in (0,\frac{1}{2}), f = 0, m_1 > 0 \\ \frac{4+3f+2m_1(1+f)}{2(4+f(2+m_1))}, & \text{if } p_m = \frac{1}{2}, f \neq 0, m_1 > 0 \end{cases} \tag{4.121}$$

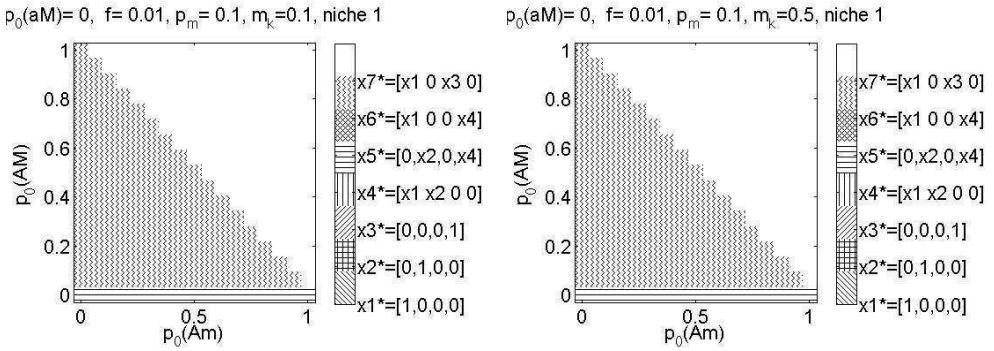
Again, it holds that $x_3^{*\text{niche 2}} = x_1^{*\text{niche 1}}$ due to the symmetric niche conditions.

The interesting case is when $p_m \neq \frac{1}{2}$ and $f \neq 0$. The value of $x_1^{*\text{niche 1}}$ depends on $m_1$ and $p_m$, and its dependence is plotted in figure 4.13. We can see that $m_1$ only has a marginal influence.



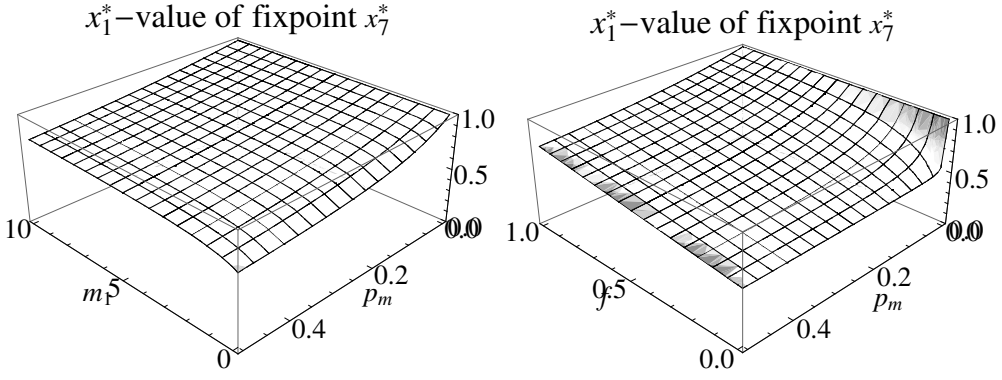

(a) $f = 0.01,\ m_1 = 0.1$ (b) $f = 0.01,\ m_1 = 0.5$

Figure 4.11: Tendency for fitness-based mating, without migration. The starting populations are set to $[x_1, x_2, 0, 1 - x_1 - x_2]$ in niche 1. The y-axis shows the initial values of gamete $AM$ and the x-axis those of $Am$. The fitness difference and mating-tendency are set to $f = 0.01$ and $m_1 = 0.1$ (left plot) and $f = 0.01$ and $m_1 = 0.5$ (right plot), and the migration rate is set to $p_m = 0$.




(a) $f = 0.01,\ m_1 = 0.1$ (b) $f = 0.01,\ m_1 = 0.5$

Figure 4.12: Tendency for fitness-based mating, with migration. The starting populations are set to $[x_1, x_2, 0, 1 - x_1 - x_2]$ in niche 1 and $[0, 1 - x_1 - x_2, x_1, x_2]$ in niche 2. The y-axis shows the initial values of gamete $AM$ and the x-axis those of $Am$. The fitness difference and mating-tendency are set to $f = 0.01$ and $m_1 = 0.1$ (left plot) and $f = 0.01$ and $m_1 = 0.5$ (right plot), and the migration rate is set to $p_m = 0.1$.




(a) $m_k = 0.5$ (b) $f = 0.01$

Figure 4.13: The $x_1^*$-value of the fixed point $\mathbf{x}_7^*$, see equation (4.121), once for $m_1 = 0.5$ (left plot) and for $f = 0.01$ (right plot).

The model presented in this section is the fitness-based mating model that I propose in this thesis. The previous sections up to this one were used to introduce step by step fitness-based mating by outlining which consequences arise from which assumptions. In the following section I will show that the haploid setting can naturally be generalized to diploid populations. The remaining of the chapter is used to analyze the consequences of fitness-based mating on the individual level and on the population level.

### 4.1.8 Diploidy

The haploid setting can easily be enlarged to a diploid setting. This is not necessary for understanding the model, furthermore, it does not change the dynamics, but for the sake of completeness I will give the generalization here.

The first locus gives rise to the genotypes $AA$, $Aa$, and $aa$ in a diploid setting. The second locus gives rise to $MM$, $Mm$, and $mm$. Let $k$ count the number of $M$-alleles at the second locus. The genotypes $MM$, $Mm$, and $mm$ correspond therefore to $k = 2$, $k = 1$, and $k = 0$, respectively. Likewise, let $i$ count the number of $A$-alleles. Let further $m_k$ denote the probability that a female with 2nd-locus genotype $k$ mates fitness-based. That means that each female with $k$ $M$-alleles mates fitness-based with a probability of $m_k$ and mates randomly with a probability of $1 - m_k$.

Instead of listing all choosing probabilities, I will give a general formula for $P_{jk}^{\text{choice}}(il)$, i.e., the probability that a $jk$-female chooses an $il$-male. Let $i^*$ denote the fittest 1st-locus genotype. Thus, for both niches, we have:

$$P_{jk}^{\text{choice}}(il) = \begin{cases} (1-m_k)P(i,l) + \delta_{i,i^*} m_k P(l|i=i^*), & \text{if } P(i^*) > 0 \\ (1-m_k)P(i,l), & \text{if } P(i^*) = 0 \end{cases} \quad (4.122)$$

with $\delta_{i,i^*}$ the Kronecker delta:

$$\delta_{i,i^*} = \begin{cases} 1, & \text{if } i = i^* \\ 0, & \text{otherwise.} \end{cases} \tag{4.123}$$

The mating probabilities $P_{\text{mate}}$ are calculated according to equation (4.98) with $P_{x_i}^{\text{accept}}(x_j) = 1$ for all $x_i, x_j$. The normalization factor for the mating probabilities then is

$$N_P = \sum_{jk} \sum_{il} P(j,k) P_{jk}^{\text{choice}}(il). \tag{4.124}$$

The simplest adjustment for the fitness values is to let them depend on the number of the fittest allele $i^*$. Thus, the homozygote $AA$ ($i = 2$) has the highest fitness value in niche 1 and the lowest fitness value in niche 2, whereas the homozygote $aa$ ($i = 0$) is the fittest in niche 2 and least the fit in niche 1. Putting this into formulae, it corresponds to

$$f^{\text{niche 1}}(ik) = 1 + i\frac{f}{2}, \quad f^{\text{niche 2}}(ik) = 1 + (2-i)\frac{f}{2} \tag{4.125}$$

or in general, for both niches:

$$f(ik) = 1 + i\frac{f}{2} + \delta_{i,i^*} f \left(1 - \frac{i}{2}\right). \tag{4.126}$$

The offspring production should naturally follow the Mendelian rules. The probability that the offspring genotype $mn$ is produced by parental genotypes $il$ and $jk$ is denoted by $P_{\text{off}}(mn|il,jk)$, as in section 2.2.1. I assume that the two loci are unlinked, thus that $P_{\text{off}}(mn|il,jk)$ is given by equation (2.3). If the two loci are linked, equation (2.3) does not hold and the explicit value of $P_{\text{off}}(mn|il,jk)$ has to be determined.

The offspring production depends on the fitness values of its parents. Let $f(il,jk)$ denote the rate at which parents $il$ and $jk$ can produce offspring such that

$$\sum_{(il,jk)} f(il,jk) P_{\text{mate}}(il,jk) = 1 \tag{4.127}$$

holds. Relative genotype frequencies in the next generation can be calculated by incorporating the appropriate terms for diploid individuals:

$$p'(mn) = \sum_{(il,jk)} f(il,jk) P_{\text{mate}}(il,jk) P_{\text{off}}(mn|il,jk). \tag{4.128}$$

Equation (4.128) is the recurrence equation with which the relative genotype frequencies in the next generation can be calculated.

### 4.1.9 Further generalizations

One possible generalization could be that fitness values are asymmetric. That means that the fitness advantage $f$ might be replaced by $f_1$ in niche 1 and by $f_2$ in niche 2. Also the migration between the niches can be asymmetric, meaning that the fraction of individuals migrating from niche 1 to niche 2 might be $p_{m_1}$ and in the other direction it might be $p_{m_2}$. A special case would be island models where $p_{m_1} > 0$ and $p_{m_2} = 0$, i.e., one niche—the island—can be reached from the other niche—the continent—but not vice versa. Of course, we would expect that the higher the $f$-value is, the higher the equilibrium frequency for $i^*$-individuals is and the higher $p_m$ is, the lower the equilibrium frequency of $i^*$ is.

## 4.2 Analysis

I have introduced the model of fitness-based mating in the previous sections (sections 4.1.1–4.1.7) by starting at simple urn models and adding successively biological details and specifications for the mating procedure. After seven such steps we arrived at the fitness-based mating model. In this section now, I will summarize and then analyze the fitness-based mating model. I will quantify the reproductive success of an individual and the influence of fitness-based mating on it. I will investigate the role of migration for the spread of fitness-based mating as well as the male mating success. This section closes with the comparison of assortative mating and fitness-based mating in terms of reproductive success of individuals expressing these preferences.

### 4.2.1 Model setup

The assumptions are as follows. The population consists of haploid individuals. Individuals have two diallelic loci. The first locus can take the alleles $A$ and $a$ and determines an ecological trait. This trait in turn determines the absolute fitness value. The second locus takes the alleles $M$ and $m$ which determine the tendency to mate fitness-based or to mate randomly.

A genotype is determined by counting the number of $A$ and $M$ alleles. So let $j \in \{0, 1\}$ denote the number of $A$-alleles at the first locus and let $k \in \{0, 1\}$ denote the number of $M$-alleles at the second locus. See table 4.10 for an overview. An individual is identified with its genotype and is denoted with $jk$. When necessary, the niche where the individual currently is in is encoded by attaching an exponent: $jk^{\text{niche 1}}$ and $jk^{\text{niche 2}}$.

Generations are non-overlapping. A generation consists of the fixed sequence of migration, couple formation, and reproduction. Mating and reproduction take place within a niche. Offspring replace parents.

**Population.** The population size is fixed. The number of elements per niche is assumed to be large enough to approximate probabilities by rel-

| | $j=0$ | $j=1$ |
|---|---|---|
| $k=0$ | $am$ | $Am$ |
| $k=1$ | $aM$ | $AM$ |

Table 4.10: The coding of the genotypes.

ative frequencies. Each niche's population consists of females and males. Each gender has the same genotype distribution as the whole niche population. Offspring are randomly assigned to each gender.

**First locus.** The first locus determines the fitness value of an individual and it is assumed to be visible as would be the performance in a mating contest. So, fitness values and 1st-locus phenotypes coincide and we can identify both. The fittest allele of a niche is denoted by $i^*$ or $j^*$. In this way, the phenotype interpretation of $i^*$, $j^*$ depends on the niche.[7]

**Second locus.** The second locus determines the mating behavior. A probability $m_k$ is assigned to the 2nd-locus genotype $k$. The value $m_k$ of a female tells us the probability with which it decides to mate a male with the fittest 1st-locus genotype. In contrast to the first locus, the 2-nd locus genotype cannot be perceived from the outside. Even though the second locus determines the mating behavior, mere observation cannot always help to infer the applied mate choice. If the mating partner is less fit, we know that the female has chosen randomly. But if the mating partner has the highest fitness value, the female could have chosen randomly or fitness-based.

**Population frequencies.** The set of all possible genotypes is denoted by $\mathfrak{A} = \{(j,k) : j,k \in \{0,1\}\}$. A population $\mathbf{p}_t$ at time $t$ is a set of elements of $\mathfrak{A}$. Elements can occur more than once in a population.

The time index $t \in \mathbb{N}_0$ denotes the generation. Generations are non-overlapping and a discrete time scale is used. Starting at $\mathbf{p}_0^n$ the model proposes how a population $\mathbf{p}_t^n$ changes to $\mathbf{p}_{t+1}^n$. Since there are two niches, there are two distinct populations $\mathbf{p}_t^{\text{niche } n}$, each containing the elements of one niche. Both niche populations influence the composition of each other in the next generation. This interdependence is caused by migration.

The relative frequency of elements having the genotype $(j,k)$ in niche $n$ is denoted by $p_t^{\text{niche } n}(j,k)$, but unless the temporal and/or geographical information is not explicitly needed, the indices are dropped such that $p_t^{\text{niche } n}(j,k)$ is shortened to $p(j,k)$. For each niche holds that $\sum_{(j,k)\in\mathfrak{A}} p(j,k) = 1$.

[7] This notation might seem to be complicating matters, but it will reduce the number of formulas without encrypting them.

The population development is modelled by formulas acting on relative genotype frequencies. These are real values and might not correspond to finite size population.

Relative allele frequencies within one niche $p(j)$ for the first locus and $p(k)$ for the second locus are calculated by

$$p(j) := \sum_{k=0}^{1} p(j,k), \quad p(k) := \sum_{j=0}^{1} p(j,k). \tag{4.129}$$

For simplicity, the notations are chosen to be as readable and intuitive as possible, leading to a slight abuse of notation: For instance, there is a semantic difference between $p(j)$ and $p(k)$, but throughout this thesis the letters $i$, $j$, $m$ are used for the first locus and $k$, $l$, $n$ for the second locus.

**Migration.** A niche determines the direction of ecological selection. A niche also represents a mating pool and matings take place only within the niche. Population elements are assigned to a niche by birth. They change to the other niche with a probability of $p_m$. Elements cannot switch niches more than once, so a element either leaves its birth niche or stays in it, but it does not switch back and forth.

The migration alters the population $\mathbf{p}_t^{\text{niche } n}$ to $\mathbf{P}_t^{\text{niche } n}$. Ignoring the time index, I will denote the relative frequency of $(j,k)$ in $\mathbf{P}^{\text{niche } n}$ with $P^{\text{niche } n}(j,k)$. Let $n$ denote one arbitrary niche and let $\bar{n}$ denote the other niche, then:

$$P^n(j,k) = p^n(j,k)(1-p_m) + \frac{N_{\bar{n}}}{N_n}\, p^{\bar{n}}(j,k)\; p_m. \tag{4.130}$$

The parameters $N_n$ and $N_{\bar{n}}$ denote the niche sizes. Assume that $N_{\bar{n}} = N_n$ so that

$$P^n(j,k) = p^n(j,k) + p_m\big(p^{\bar{n}}(j,k) - p^n(j,k)\big) \tag{4.131}$$

holds.

**Fitness values.** The fitness value, see section 4.1.3 how the term "fitness" is used in this thesis, of an individual is determined by its first locus. I assume that the first locus codes a trait that tells us the performance in a mating contest. The performance depends on the niche only, so that the fitness value does not include sexual selection. The fitness value of an individual is its expected offspring number from one single mating. The overall number of offspring of the focal individual also depends on its mating decisions and the number of matings. For that reason, I distinguish between *fitness* and *reproductive success*.

Selection favors $A$-carriers in niche 1 and $a$-carriers in niche 2. Fitness values are denoted by $f(j,k) = f(j)$ and have been defined as

$$f(j) = 1 + f\delta_{j,j^*}, \tag{4.132}$$

| 1st-locus genotype | $A$ | $a$ |
|---|---|---|
| niche 1 | $1+f$ | 1 |
| niche 2 | 1 | $1+f$ |

Table 4.11: Fitness values.

i.e., as in table 4.11. Fitness becomes relevant during the mating procedure after migration. The fitness of a couple $(i,l)$ and $(j,k)$, that is, its expected number of offspring is denoted by $F(il,jk) = F(i,j)$ with $F(i,j)$ being the average fitness of $(i,l)$ and $(j,k)$:

$$F(i,j) = \frac{f(i)+f(j)}{2} = 1 + \frac{f}{2}\big(\delta_{j,j^*} + \delta_{i,i^*}\big). \tag{4.133}$$

The mean couple fitness $F$ is calculated using the choosing probabilities $P^{\text{choice}}$ and has been used already in section 4.1 for the normalization of subsequent genotype frequencies:

$$F = \sum_{il,jk} P(j,k) P^{\text{choice}}_{jk}(il) F(i,j). \tag{4.134}$$

**Mating procedure.** The population consists of females and males. Both genders have the same gametic distribution and therefore, they do not have to be distinguished. Each female chooses a partner and then mates. Males can be chosen more than once, therefore, the population is monoandrous and polygynous. A female chooses on grounds of her mate preference that is determined by her 2nd-locus allele. She either mates fitness-based (with a probability of $m_k$) or randomly (with a probability of $1-m_k$). If she mates fitness-based, then the male is chosen among the $i^*$-carriers. She does not mate, when there are no $i^*$-carriers. If she mates randomly, her mate is chosen from the whole set of males. Thus, the probability that a $jk$-female chooses an $il$-male is:

$$P^{\text{choice}}_{jk}(il) = \begin{cases} (1-m_k)P(i,l) + \delta_{i,i^*} m_k P(l|i=i^*), & \text{if } P(i^*) > 0 \\ (1-m_k)P(i,l), & \text{if } P(i^*) = 0. \end{cases} \tag{4.135}$$

Males accept any mating offer.

The loci are assumed to be situated on different chromosomes, such that the recombination rate between both loci can be set to 1/2.

### 4.2.2 Analysis

**Recurrence equation.** The general recurrence equation for the relative genotype frequency of gamete $(o,p)$ can be derived:

$$p_{t+1}(o,p) = \sum_{(j,k)\in\mathfrak{A}} P_t(j,k) \sum_{(i,l)\in\mathfrak{A}} P_{\text{off}}(op|il,jk) P^{\text{choice}}_{jk}(il) \frac{F(i,j)}{F}. \tag{4.136}$$

Equation (4.136) tells us the genotype distribution of the subsequent generation when the current distribution is given. It will be iterated computationally in chapter 5.

**Reproductive success.** The second locus determines the mating behavior and is not subject to ecological selection. Hence, its influence can only be determined by checking how $m_k$ affects the reproductive success. The reproductive success of $(j,k)$ in a given niche is denoted by $\phi(j,k)$. Note the difference to $f(j,k)$ which gives the genetically determined fitness. The fitness value does not depend on the population composition. The reproductive success, in contrast, includes the effects of mating behavior, population composition, migration effects, and ecological performance. I will quantify the reproductive success of a $(j,k)$-female as

$$\phi(j,k) = \sum_{(i,l)\in\mathfrak{A}} P_{jk}^{\text{choice}}(il)F(i,j) \tag{4.137}$$

$$= \begin{cases} 1+\frac{f}{2}\Big(\delta_{j,j^*}+(1-m_k)P(i^*)+m_k\Big), & \text{if } P(i^*)>0 \\ 1-m_k, & \text{if } P(i^*)=0. \end{cases} \tag{4.138}$$

The mean couple fitness $F$ in a given niche includes the population composition and the mate preferences, and is the same as the mean reproductive success:

$$F = \sum_{j,k}\phi(j,k)P(j,k) \tag{4.139}$$

$$= \begin{cases} 1+fP(i^*)+\frac{f}{2}\bar{m}\big(1-P(i^*)\big), & \text{if } P(i^*)>0 \\ 1-\bar{m}, & \text{if } P(i^*)=0, \end{cases} \tag{4.140}$$

with $\bar{m}=\sum_k m_k P(k)$.

**Effect on the reproductive success.** If fitness-based mating is beneficial for a female $(j,k)$, it increases its reproductive success $\phi(j,k)$. We are interested whether $\phi(j,k)$ increases with the tendency to mate fitness-based. Let us first look at the derivative of $\phi$ with respect to $m_k$:

$$\frac{\partial\phi(j,k)}{\partial m_k} \overset{(4.138)}{=} \begin{cases} \frac{f}{2}(1-P(i^*)), & \text{if } P(i^*)>0 \\ -1, & \text{if } P(i^*)=0. \end{cases} \tag{4.141}$$

The reproductive success is constant when $P(i^*)=1$ and it increases with $m_k$ as long as $P(i^*)\in(0,1)$. Of course, when $i^*$ is missing, it does not pay to mate fitness-based on the contrary, the female forgoes mating chances. Let us now look at the derivative of $F$ with respect to $m_k$:

$$\frac{\partial F}{\partial m_k} \overset{(4.140)}{=} \begin{cases} P(k)\frac{f}{2}\big(1-P(i^*)\big), & \text{if } P(i^*)>0 \\ -P(k), & \text{if } P(i^*)=0. \end{cases} \tag{4.142}$$

Also, $\partial F/\partial m_k$ is zero when $P(i^*)=1$ and it is positive as long as $P(i^*)\in(0,1)$. So, both the individual reproductive success and the mean couple fitness increase with $m_k$.

**Effect on the population composition.** The term $P(j,k)\phi(j,k)/F$ gives a measure of the proportion of $(j,k)$-offspring in the next generation. Of course, offspring do not necessarily have the same genotype as their parents. It is rather determined by Mendelian inheritance laws. Furthermore, the term captures the offspring of $(j,k)$-females only. Please look at the derivative of the offspring proportion with respect to $m_k$:

$$\frac{\partial}{\partial m_k} P(j,k)\frac{\phi(j,k)}{F} = \frac{P(j,k)}{F^2}\left(F\frac{\partial \phi(j,k)}{\partial m_k} - \phi(j,k)\frac{\partial F}{\partial m_k}\right) \tag{4.143}$$

The term

$$F\frac{\partial \phi}{\partial m_k} - \phi\frac{\partial F}{\partial m_k} \tag{4.144}$$

is decisive for the sign of equation (4.143), because the other factor is non-negative. There are three cases. The first case treats $P(i^*) = 0$:

$$F\frac{\partial \phi}{\partial m_k} - \phi\frac{\partial F}{\partial m_k} = \bar{m} - m_k P(k) - \big(1 - P(k)\big) \tag{4.145}$$

$$= \begin{cases} m_1 P(l=1) + P(l=0) - 1, & \text{if } k = 0 \\ m_0 P(l=0) + P(l=1) - 1, & \text{if } k = 1. \end{cases} \tag{4.146}$$

Since $m_0, m_1 \in \mathbb{R}[0,1]$ the terms of equation (4.146) range in $\mathbb{R}[P(l=0)-1, 0]$ for $k=0$ and $\mathbb{R}[P(l=1)-1, 0]$ for $k=1$. Thus, the term (4.144) is smaller or equal to zero. As expected, the relative frequency of $(j,k)$-offspring decreases when $(j,k)$ mates with a higher probability fitness-based when there is no $i^*$-male present in the population. In that case an $jk$-female forgoes even more mating chances.

The second and third case both treat $P(i^*) > 0$ and distinguish $\delta_{j,j^*} = 0$ (2nd case) and $\delta_{j,j^*} = 1$ (3rd case). For the second case, it holds that

$$\begin{aligned} F\frac{\partial \phi}{\partial m_k} - \phi\frac{\partial F}{\partial m_k} &= \frac{f}{2}\big(1 - P(i^*)\big) * \dots \\ &\quad * \left[1 - P(k) + fP(i^*)\left(1 - \frac{1}{2}P(k)\right) + \dots \right. \\ &\quad \left. + \frac{f}{2}\big(1 - P(i^*)\big)\big(\bar{m} - m_k P(k)\big)\right] \\ &\geq 0. \end{aligned} \tag{4.147}$$

Hence, the fraction of $(j,k)$-offspring increases with $m_k$ as long as $P(i^*) < 0$. As expected, a less fit female profits from mating fitness-based with higher $m_k$.

In the third and last case, when $P(i^*) > 0$ and $\delta_{j,j^*} = 1$, it holds that

$$\begin{aligned} F\frac{\partial \phi}{\partial m_k} - \phi\frac{\partial F}{\partial m_k} &= \frac{f}{2}\big(1 - P(i^*)\big) * \dots \\ &\quad * \left[1 - P(k)\left(1 + \frac{f}{2}\right) + fP(i^*)\left(1 - \frac{1}{2}P(k)\right) + \dots \right. \\ &\quad \left. + \frac{f}{2}\big(1 - P(i^*)\big)\big(\bar{m} - m_k P(k)\big)\right]. \end{aligned} \tag{4.148}$$

Here, the sign can become negative. The expected increase in reproductive success of $i^*$-females may be relatively weaker than that of less fit females when the latter are in the minority, $f$ is large, and $P(k)$ is high. Nevertheless, this does lead to an increase of $i^*$ in the next generation, as we will see when we compute the effect on $p'(i \neq i^*)$ and $p'(i^*)$ when an $i^*$-female switches from a randomly chosen mate to a mate chosen fitness-based. This switch is equivalent to an increases of $m_k$. Let $P^{\text{fbm}}_{x_i}(x_j)$ and $P^{\text{rm}}_{x_i}(x_j)$ denote the non-normalized frequency of an offspring with type $x_j$ produced by an $x_i$-female under fitness-based mating and random mating, respectively. Let us concentrate on niche 1, where $i^* \hat{=} A$. We then know that

$$
\begin{aligned}
P^{\text{fbm}}_A(A) &= 1+f && (4.149)\\
P^{\text{fbm}}_A(a) &= 0 && (4.150)\\
P^{\text{fbm}}_a(A) &= \frac{1}{2}\left(1+\frac{f}{2}\right) && (4.151)\\
P^{\text{fbm}}_a(a) &= \frac{1}{2}\left(1+\frac{f}{2}\right) && (4.152)\\
P^{\text{rm}}_A(A) &= P(a)\frac{1}{2}\left(1+\frac{f}{2}\right)+P(A)(1+f) && (4.153)\\
P^{\text{rm}}_A(a) &= P(a)\frac{1}{2}\left(1+\frac{f}{2}\right) && (4.154)\\
P^{\text{rm}}_a(A) &= P(A)\frac{1}{2}\left(1+\frac{f}{2}\right) && (4.155)\\
P^{\text{rm}}_a(a) &= P(a)+P(A)\frac{1}{2}\left(1+\frac{f}{2}\right). && (4.156)
\end{aligned}
$$

The frequency difference in the next generation of offspring of type $A$ and $a$ through their mother's switch from random mating to fitness-based mating is then

$$
\begin{aligned}
\Delta p'(A) &:= p'^{\text{fbm}}(A) - p'^{\text{rm}}(A) && (4.157)\\
&= P(A)[P^{\text{fbm}}_A(A) - P^{\text{rm}}_A(A)] + P(a)[P^{\text{fbm}}_a(A) - P^{\text{rm}}_a(A)] && (4.158)\\
&= P(a)\left[\frac{1}{2}\left(1+\frac{f}{2}\right)P(a) + \frac{1}{2}\left(1+\frac{3f}{2}\right)P(A)\right] && (4.159)
\end{aligned}
$$

and likewise is

$$
\begin{aligned}
\Delta p'(a) &:= p'^{\text{fbm}}(a) - p'^{\text{rm}}(a) && (4.160)\\
&= P(a)\left[\frac{1}{2}\left(\frac{f}{2}-1\right)P(a) - \frac{1}{2}\left(1+\frac{f}{2}\right)p(A)\right]. && (4.161)
\end{aligned}
$$

Clearly,

$$
\Delta p'(A) > \Delta p'(a) \qquad (4.162)
$$

holds. So even though, the expected increase in the reproductive success of less fit females may be relatively stronger, the frequency of $i^*$-offspring increases with $m_k$. Furthermore, the less fit females are no longer breeding true after switching, and less fit males loose their mates. Thus, fitness-based mating will make the selective advantage of $i^*$-carriers even stronger.

**Migration effect.** This paragraph examines the effect of migration on the individual reproductive success. The migration affects $P(j,k)$. Again let $n$ denote one arbitrary niche and $\bar{n}$ the other niche, then:

$$\frac{\partial P^n(j,k)}{\partial p_m} \overset{(4.131)}{=} p^{\bar{n}}(j,k) - p^n(j,k). \tag{4.163}$$

Let $i_n^*$ be the fittest 1st-locus allele in niche $n$. Then

$$\frac{\partial \phi^n(j,k)}{\partial p_m} = (1-m_k)\frac{f}{2}\Big(p^{\bar{n}}(i_n^*) - p^n(i_n^*)\Big). \tag{4.164}$$

If the relative frequency of $i_n^*$ is higher in niche $\bar{n}$, where $i^*$ is not fittest, than in niche $n$, then migration helps increasing $i^*$ in the "right niches", because the immigration of $i^*$ is higher than its emigration. In this case, $\partial\phi^n/\partial p_m > 0$ holds. If $p^n(i_n^*)$ is higher than $p^{\bar{n}}(i_n^*)$, migration counteracts the selection's effort to increase the relative frequency of $i_n^*$ in niche $n$. And then it holds that $\partial\phi^n/\partial p_m < 0$.

Hence, if selection has resulted in a higher proportion of $i^*$-genotypes in the appropriate niches, then also $\phi^n(j,k)$ is negatively influenced by a higher migration rate $p_m$. Of course, in a non-degenerate case, we find that $p^n(i_n^*) \geq p^{\bar{n}}(i_n^*)$.

An increase of $m_k$ increases the individual's reproductive success when $P(i^*) \in (0,1)$, see equation (4.141). But the slope of the reproductive-success increase depends on the migration rate, because

$$\frac{\partial^2 \phi^n(j,k)}{\partial p_m \partial m_k} = -\frac{f}{2}\Big(p^{\bar{n}}(i_n^*) - p^n(i_n^*)\Big). \tag{4.165}$$

In situations where $p^n(i_n^*) > p^{\bar{n}}(i_n^*)$, which is to be expected under non-degenerate conditions, the slope of the reproductive-success increase is positive. That means that the bigger the migration rate is, the more a female profits from fitness-based mating. Thus, while the overall reproductive success decreases with higher migration, the relative advantage of fitness-based mating increases. Therefore, high $m_k$-values dampen the migration effect in non-degenerate cases. Just like selection, fitness-based mating acts contrary to migration. We will find this confirmed in section 5.2.2 where fitness-based mating maintains $i^*$ at higher frequencies than any other mating scheme tested.

**Probabilities of being selected for mating.** This paragraph concentrates on the males' chances of being chosen. The probability $P_{\text{chosen}}(i,l)$, with

which an $il$-male is chosen for mating, is calculated by

$$P_{\text{chosen}}(i,l) = \sum_{(j,k)\in\mathfrak{A}} P_{jk}^{\text{choice}}(il)P(j,k) \tag{4.166}$$

$$= P(il)(1-\bar{m}) + \delta_{i,i^*}P(l|i=i^*)\bar{m}, \tag{4.167}$$

with $\bar{m} = \sum_k m_k P(k)$. Summing equation (4.167) over $l$ gives

$$P_{\text{chosen}}(i) = (1-\bar{m})P(i) + \delta_{i,i^*}\bar{m} = P(i) + \bar{m}\big(\delta_{i,i^*} - P(i^*)\big). \tag{4.168}$$

Apart from its own relative frequency, $i^*$-males have an additional term in $P_{\text{chosen}}$. As expected from the model setup, the probability of being selected as a mating partner is increased for $i^*$-males because of the existence of fitness-based mating. Existence of fitness-based mating in the population means that $\bar{m} > 0$, and this in turn means that there exists a $k$ with $P(k) > 0$ and $m_k > 0$.

Summing equation (4.167) over $i$ gives $P_{\text{chosen}}(l)$ which is the probability that an $l$-carrier is chosen:

$$P_{\text{chosen}}(l) = P(l) + \bar{m}\Big(P(l|i=i^*) - P(l)\Big). \tag{4.169}$$

The dependence of $P_{\text{chosen}}$ on the 1st-locus genotype, see equation (4.168), is not surprising, since the model enables mate preferences which are based on the 1st-locus genotype. But equation (4.169) shows that the probability for the second locus of being chosen can deviate from the respective probability under random mating. In a random-mating situation, the probability of a genotype of being chosen would simply be its relative frequency in the population. Now in contrast, the probability is lowered by fitness-based mating when $P(l|i^*) < P(l)$ and is increased when $P(l|i^*) > P(l)$. Thus, the 2nd-locus allele which co-occurs more often with $i^*$ in a gamete will have higher chances of being selected than under random mating. Presumably, the 2nd-locus allele $k$ with highest $m_k$ will get linked to $j^*$, because it mates preferentially with $j^*$ and because it passes $k$ as well as $j^*$ to its offspring. I will monitor $P_{\text{chosen}}$ in computer simulations in section 5.2.7 and will confirm that $M$ co-occurs more often with $i^*$ than $m$ does.

### 4.2.3 Fitness-based mating in Gavrilets' framework

Gavrilets [29] (pp. 280–287) has proposed a general framework in which non-random mating can be modelled. He has proposed mating pools in which all individuals can potentially mate with each other. Individuals from different mating pools do not meet and hence do not mate.

Gavrilets distinguishes two cases. In the first case, individuals join preferentially a mating pool and mate randomly within this pool. In the second case, individuals join randomly a mating pool and mate preferentially. In the latter case, encounters are random but matings depend on mate preferences. The fitness-based mating model falls into the second case where each niche's population forms one mating pool in which individuals mate preferentially.

In Gavrilets' framework, the preference of $jk$-females for $il$-males is denoted by $\psi_{(j,k),(i,l)}$ and depends for fitness-based mating only on the female preference $k$ and the male's 1st-locus genotype $i$. Under fitness-based mating it is

$$\psi_{(j,k),(i,l)} = \begin{cases} 1 - m_k, & \text{if } j \neq j^* \\ 1 + m_k \frac{1-P(j^*)}{P(j^*)}, & \text{if } j = j^*. \end{cases} \tag{4.170}$$

Gavrilets introduced a function $f(\bar{\psi})$ that depends on the average strength $\bar{\psi}$ of the preference:

$$\bar{\psi}_{(j,k)} = \sum_{(i,l)} \psi_{(j,k),(i,l)} P(i,l) = 1 \tag{4.171}$$

and gives the probability that a female $(j,k)$ does mate at all. Since $\bar{\psi}_{(j,k)} = 1$, i.e., each female mates, the function is constant and all terms involving $f(\bar{\psi})$ cancel out in the equations.

Gavrilets defines the relative mating success of females, $w_{f,jk}$, and males, $w_{m,il}$, by

$$w_{f,jk} = \frac{\sum_{(i,l)} P(jk \times il)}{P(jk)}, \quad w_{m,il} = \frac{\sum_{(j,k)} P(jk \times il)}{P(il)} \tag{4.172}$$

with $P(jk \times il)$ being the mating probability of a $jk$-female and an $il$-male. Using the notation of this thesis $P(jk \times il)$ amounts to

$$P(jk \times il) = \begin{cases} (1 - m_k) P(il) P(jk), & \text{if } i \neq i^* \\ \left(1 + m_k \frac{1-P(i^*)}{P(i^*)}\right) P(i^* l) P(jk), & \text{if } i = i^*, \end{cases} \tag{4.173}$$

and this can be written as [29]:

$$P(jk \times il) = \psi_{(j,k),(i,l)} P(j,k) P(i,l). \tag{4.174}$$

Using (4.172) and (4.173), the relative mating success for females and males then are

$$w_{f,jk} = 1, \quad w_{m,il} = \begin{cases} 1 - \bar{m}, & \text{if } i \neq i^* \\ 1 + \bar{m} \frac{1-P(i^*)}{P(i^*)}, & \text{if } i = i^*. \end{cases} \tag{4.175}$$

The mate preference is non-selective for females, which means that their choosiness is not costly even when the preferred mate is rare. In contrast to females, the mate preferences is selective for males, giving advantages to $i^*$-carriers.

Gavrilets distinguishes similarity-based and matching-based mate preferences. If *"mating is controlled by a single trait [...] expressed in both*

*sexes"* (e.g., as in assortative mating), he speaks of similarity-based mating. If *"mating is controlled [...] by two different sex-linked traits"*, it is matching-based. According to these definitions, fitness-based mating is neither similarity-based nor matching-based. Clearly, fitness-based mating is not similarity-based, because $M$-carriers do not care whether mates have also an $M$-allele. But fitness-based mating is also not matching-based, because matching-based models work as follows: Mating trait $A$ prefers trait $B$, and mating trait $a$ prefers trait $b$. That means, different preferential maters prefer different traits, and their genotype at the mating locus determines which trait they prefer. But in the fitness-based mating model, the preferred trait depends on the niche only, and the genotype at the mating locus is the same for all preferential maters.

### 4.2.4 The reproductive success for fitness-based and assortative mating

In this section I will examine how the different mating rules manifest themselves in the reproductive success of an individual and I will examine which mating rule is more beneficial.

Let $\phi_{\text{fbm}}$ denote the reproductive success under fitness-based mating and let $\phi_{\text{am}}$ denote the reproductive success under assortative mating. For a $jk$-female, $\phi_{\text{fbm}}(j,k)$ is given in the equations (4.137)–(4.138). For the calculation of the reproductive success under assortative mating, we need the probability that $(j,k)$ chooses $(i,l)$. Analogously to equation (4.135), that probability is

$$P_{jk}^{\text{choice,am}}(il) = (1-n_k)P(i,l) + \delta_{i,j}n_kP(l|i=j) \tag{4.176}$$

where $n_k$ denotes the tendency of $k$-carriers to mate assortatively. Thus,

$$\phi_{\text{am}}(j,k) \;=\; 1 + \frac{f}{2}\big((1+n_k)\delta_{j,j^*} + (1-n_k)P(i^*)\big). \tag{4.177}$$

Assume that the tendencies to mate assortatively or to mate fitness-based are the same: $n_k = m_k$. The difference in the reproductive success then is

$$\Delta\phi(j,k) \;=\; \phi_{\text{fbm}}(j,k) - \phi_{\text{am}}(j,k) \tag{4.178}$$

$$\overset{(4.138),(4.177)}{=} \begin{cases} \frac{f}{2}n_k(1-\delta_{j,j^*}) \geq 0, & \text{if } P(i^*)>0 \\ -m_k, & \text{if } P(i^*=0). \end{cases} \tag{4.179}$$

For $j^*$-females, both mate preferences perform equally well, because the fittest males are also similar to them. Fitness-based mating is clearly more profitable for less fit females when $i^*$-males are present. Additionally, the proportions of offspring with genotype $i^*$ are higher when mating fitness-based, because one parent is of genotype $i^*$. However, if $P(i^*)=0$, then fitness-based maters forgo mating chances while assortative maters do not. Clearly, in this situation, assortative mating is more profitable.

## 4.3 Summary

In this chapter, I have proposed a population-genetic model for fitness-based mating. Fitness-based mating is a mate preference where the preference is linked to the absolute fitness value of potential mates. The model has been formalized in mathematical terms. The number of two loci is minimal, such that analytical treatment is feasible.

In the first part of this chapter, the fitness-based mating model has been introduced step by step by demonstrating the effect of each single assumption. The second part has treated the analysis of fitness-based mating; I will highlight the results:

I have derived the explicit solution of the gametic equilibrium frequencies, see equation (4.121). Additionally, the recurrence equation giving the genotype distribution of the subsequent generation when the current distribution is known has been derived, see equation (4.136). It will be iterated for computing the equilibrium distribution in chapter 5.

The reproductive success, $\phi(j,k)$, and the mean couple fitness, $F$, increase with higher $m_k$ under the condition that $P(i^*) \in (0,1)$. This should be confirmed by the fixation of that 2nd-locus genotype $k$ which has the highest $m_k$, see equations (4.141)–(4.142).[8]

The influence of fitness-based mating on the population composition has been investigated. The fraction of $jk$-offspring can be increased by a higher $m_k$, that means by a stronger preference when the preferred partner is around, see equations (4.147) and (4.162).

The interplay between migration and fitness-based mating has been examined. The higher $p_m$, the higher the selection pressure for fitness-based mating, see equation (4.165).

The mating probabilities of the second locus have been investigated. It turns out that the fixation of fitness-based mating is not only due to its increasing effect on the reproductive success, but is assisted by non-random mating probabilities. Simulations, which can be found in chapter 5, confirm that this is due to the association of genes for fitness-based mating with $i^*$.

A translation of fitness-based mating into Gavrilets' framework for non-random mating [29] proved that fitness-based mating is non-selective for females and selective for males. This means that the choosiness is not costly even when the preferred trait is rare, see equation (4.175).

Fitness-based mating is proposed as an alternative mating strategy to assortative mating. The effects on the individual reproductive success of both strategies have been compared. In this framework, fitness-based mating and assortative mating perform equally well for fit females. But fitness-based mating is superior to assortative mating for less fit females, see equation (4.179). However, if no $i^*$-individual is present, fitness-based maters forgo mating chances, and assortative mating is more profitable.

[8] In diploid populations, the condition that "given that $k$ is homozygotic" has to be added, because heterozygotes cannot go to fixation according to the Mendelian inheritance rules.

Both mating strategies coincide in the trivial case that the population consists of $i^*$-individuals only. This gives rise to the conjecture that if both mating schemes would compete in one population, then fitness-based mating would be more successful. Condition is that the population is distributed across two niches with opposite selection pressures. In this case, the relative frequency of $i^*$ remains between 0 and 1, and fitness-based mating would be more profitable than assortative mating.

# 5 Simulation results

This chapter presents the results of three different mating schemes which are implemented into computer models: fitness-based mating, assortative mating, and random mating. The first two were chosen to complement the comparison between fitness-based and assortative mating carried out earlier in this thesis (biologically: section 2.4, analytically: section 4.2.4). I have chosen to contrast their dynamics with the dynamics resulting when there is no mate preference, i.e., random mating.

The chapter is separated into three parts: Section 5.1 presents implementation details and predictions. Section 5.2 lists the results of iterating the recurrence equations for all three mating models (fitness-based, assortative, random mating), while section 5.3 presents several finite-size implementations of fitness-based mating only.

## 5.1 Preparation

### 5.1.1 Recurrence equations for fitness-based, assortative, and random mating

The equilibrium distribution for the fitness-based mating model has been calculated in equation (4.121). Also, the equilibrium under random mating is known from section 4.1 as I will show later in section 5.1.2. However, I did not provide the equilibrium distribution under assortative mating. And so I will iterate the recurrence equation instead of plotting the equilibrium solutions for various parameters. I will rather use the iteration method, because we then obtain the equilibrium distribution for all three mating schemes and we are also able to monitor the transient behavior of the populations approaching the equilibrium. Furthermore, we can compare in this way the time needed for approaching the equilibrium.

Let $\hat{P}$ denote the relative equilibrium frequency, then equation (4.136) is used to obtain the equilibrium equation for gamete *op* in each niche:

$$\hat{P}(o,p) = \sum_{(j,k)} \sum_{(i,l)} \hat{P}(j,k) P_{\text{off}}(op|il,jk) P_{jk}^{\text{choice}}(il) \frac{F(i,j)}{\bar{F}}. \tag{5.1}$$

The three mating models differ in the choosing probabilities $P_{jk}^{\text{choice}}(il)$. The fitness-based mating model requires $P_{jk}^{\text{choice}}(il)$, as given in equation (4.135)

$$P_{jk}^{\text{choice}}(il) = \begin{cases} (1-m_k)P(i,l) + \delta_{i,i^*} m_k P(l|i=i^*), & \text{if } P(i^*) > 0 \\ (1-m_k)P(i,l), & \text{if } P(i^*) = 0. \end{cases} \tag{5.2}$$

The assortative model uses $P_{jk}^{\text{choice}}(il)$, as given in equation (4.176)

$$P_{jk}^{\text{choice}}(il) = (1 - n_k)P(i,l) + \delta_{i,j} n_k P(l|i = j). \tag{5.3}$$

And the model with random mating demands:

$$P_{jk}^{\text{choice}}(il) = P(i,l). \tag{5.4}$$

All three equations (5.2)–(5.4) coincide with each other for $m_k = n_k = 0$.

### 5.1.2 Predictions

**Predictions for fitness-based mating.** The reproductive success $\phi^n$ of genotype $(j,k)$ was given by equation (4.138) and its dependence on the $m_k$, i.e., the tendency to mate fitness-based, by equation (4.141). If $P(i^*) \in (0,1)$, genotypes with higher $m_k$ have a higher reproductive success than genotypes with lower $m_k$. As long as both niches are linked by migration, $P(i^*)$ will stay below 1 and we know from section 4.1.7 that fitness-based mating spreads then in both niches and that allele $M$ goes to fixation.

The change in the relative advantage of fitness-based mating with respect to the migration rate is given by equation (4.165) and is positive when $p^n(i_n^*) > p^{\bar{n}}(i_n^*)$. A natural, or non-degenerate, situation reflects the work of natural selection in favor of $i^*$ in the respective niches. Therefore, we can assume $p^n(i_n^*) > p^{\bar{n}}(i_n^*)$. And since the inequality is presumably met, the pressure to mate fitness-based should be the higher, the higher the migration rate $p_m$ is. As a consequence, the higher $p_m$ is, the earlier should be the point in time of the fixation of $M$.

Fitness-based mating strengthens the ecological selection pressure by increasing the the mating chances of fit males. We have seen this at equation (4.121) and figure 4.13 where the value of $i^*$ increases with $m_k$.

**Predictions for assortative mating.** The reproductive success of a $jk$-female is given by equation (4.177) and its derivate with respect to $n_k$ is

$$\frac{\partial}{\partial n_k}\phi(j,k) = \frac{f}{2}\big(\delta_{j,j^*} - P(i^*)\big) = \begin{cases} -\frac{f}{2}P(i^*) & \text{if } j \neq j^* \\ \frac{f}{2}\big(1 - P(i^*)\big) & \text{if } j = j^*. \end{cases} \tag{5.5}$$

Clearly, for less fit individuals, it is better to mate randomly, because they otherwise choose a partner that has equally low fitness. Fit individuals profit from increasing their tendency to mate assortatively as long as $P(i^*) < 1$. Since a part of the population does not profit from assortative mating, it is not clear whether the mate preference goes to fixation. But if it does, we can expect that the fixation is reached slower when compared to fitness-based mating populations.

**Predictions for random mating.** In the absence of mate preferences, selection and migration work as the only counteracting evolutionary operators. Therefore, the relative equilibrium frequency for allele $A$ should

resemble the known selection-migration-equilibrium as given in equation (4.96). However, the difference is that $x_2^*$ now resembles the relative frequency of allele $A$ in niche 1 instead that of gamete $Am$. Since the mating alleles, $M$ and $m$, have no effect, both remain in the population. The 1st-locus alleles, $A$ and $a$, make a difference for the equilibrium frequency of a gamete. That means for niche 1 that $P(AM) = P(Am) = x_A^*/2$ and $P(aM) = P(am) = (1 - x_A^*)/2$ with $x_A^*$ given as $x_2^*$ in equation (4.96).

### 5.1.3 Notes on the computer implementation

**Precision.** The numerical precision depends on the precision supported by the software and computer architecture. All following simulations were conducted with *MATLAB*® and have been run on a *Sun Fire 880*-machine equipped with a *Sparc US-III+*, 64-bit processor. The operation system was *SUN OS 5.10* and the *MATLAB*® version was *7.3.0.298 (R2006b)*.

**Parameters.** The parameters are the migration rate $p_m$, the fitness difference $f$, the mating tendencies $m_k$, and the initial populations. I will focus on the parameter space spanned by $p_m \in \mathbb{R}[0, 0.5]$, $f \in \mathbb{R}[0, 2]$ (partially also $f \in \mathbb{R}[0, 4]$), and $m_1, n_1 \in \mathbb{R}[0, 1]$. I will set $m_0 = n_0 = 0$, because fitness-based or assortative mating shall compete with random mating. Clearly, the higher $m_1$ or $n_1$ are, the higher the strength of sexual selection is.

With the exception of parameter $f$, the entire parameter range is examined. The maximal value for the migration rate is 0.5. This value marks the situation where individuals arbitrarily choose the niches where they live in. Above $p_m = 0.5$, there is no habitat bonding, which means that individuals rather leave their birth niche. Preferring the fittest partner of the current niche does not necessarily equip offspring with genes adapted to the niche they wander off to. Thus, the range of $p_m > 0.5$ is not relevant for fitness-based mating.

Theoretically, the parameter $f$ can take any value in $\mathbb{R}_0^+$, and covering the whole range is, of course, impractical. Therefore, I restrict the examinations to cases when $f$ is rather low which refers to situations when the ecological selection is marginal. However, I have derived in equation (4.121) the explicit equilibrium frequency of $AM$ in niche 1 and $aM$ in niche 2. So, the equilibrium distribution under any other combination of the parameters can be obtained by plugging the parameters into the equation (4.121).

The iteration of the recurrence equations will start from homogeneous initial populations. So, each gamete will form 25% of each niche's population. I have chosen this starting population, because such a homogeneous population would result from a randomly mating population where each allele has equal frequency and selection pressure is absent. But as I have outlined in section 4.1.7, all starting populations—except those where more than one gamete are initially absent—lead to the same equilibrium, i.e., the fixation of allele $M$.

## 5.2 Results

### 5.2.1 Iterating the recurrence equations

Figure 5.1 shows the results of iterating the recurrence equations starting from a uniform population for (a) fitness-based, (b) assortative, and (c) random mating. The development of niche 1 is plotted only, because the start conditions are the same and the niche conditions are symmetric, such that the evolution in all three models also shows a symmetric behavior.

Allele $M$ goes to fixation for fitness-based mating and assortative mating, confirming that both mating strategies convey an evolutionary advantage over random mating (allele $m$). In the "control" model of exclusively random mating individuals, both alleles, $M$ and $m$, remain in the population at equal frequencies.

Due to natural and sexual selection, the fittest 1st-locus allele ($A$ in niche 1, and $a$ in niche 2) dominates the population in all three models. The relative frequency difference of $AM$ (of $aM$ in niche 2) to the less fit gamete ($aM$ in niche 1, and $AM$ in niche 2) is highest under fitness-based mating. Under assortative and random mating, the frequency difference between fit and less fit gametes is quite small (note that under random mating, two equally fit gametes, namely $AM$ and $Am$ in niche 1, persist in the population). Thus, fitness-based mating permits the highest equilibrium frequency of fit gametes.

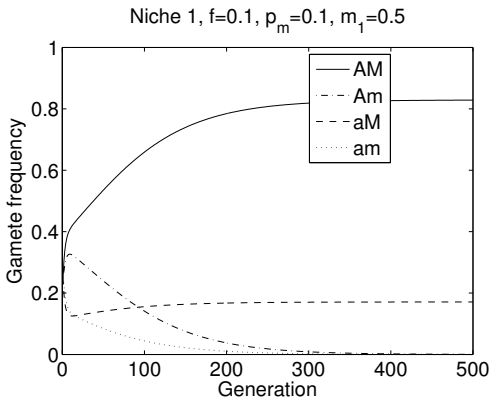
In none of the three models, a linkage disequilibrium in the equilibrium population persists. For fitness-based and assortative mating, this is due to the absence of allele $m$. For random mating, the population does not develop a linkage disequilibrium, because the second locus is irrelevant.

Populations with random mating reach the equilibrium fastest, populations with fitness-based mating reach the equilibrium second-fastest. Populations with assortative mating require the most iterations.

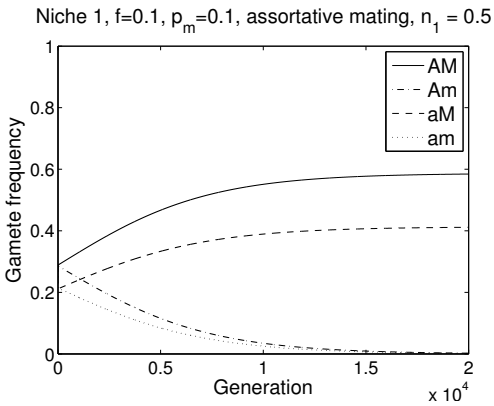
### 5.2.2 Migration rate

Migration introduces a continual influx of less fit individuals, because they immigrate from a niche where an opposite selection scheme reigns. Thus, the equilibrium frequency of $i^*$ should decrease with $p_m$. Figure 5.2 shows the relative equilibrium frequencies for various migration rates $p_m$ and confirms this expectation. The simulation results show that selection is strong enough to outweigh the mixing effect of even strong migration for all three mating schemes.

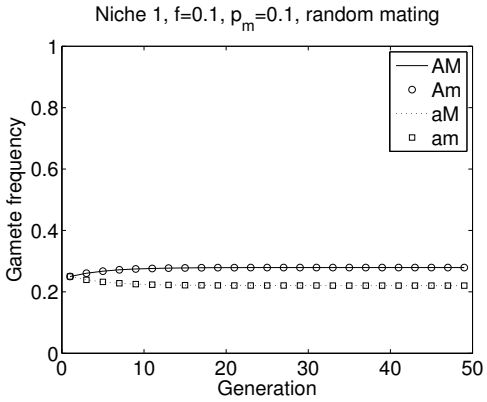
Going into detail, the relative equilibrium frequency of $AM$ decreases with $p_m$, whereas $aM$ increases thus, the strength of the 1st-locus polymorphism decreases. In contrast to randomly mating populations, the relative frequency of $m$ remains at zero under fitness-based and assortative mating for low to moderate migration rates. But if migration rates are high, assortative mating and random mating maintain $m$ at a comparable relative frequency. In this way, only fitness-based mating proliferates




(a) Fitness-based mating ($m_0 = 0$, $m_1 = 0.5$)

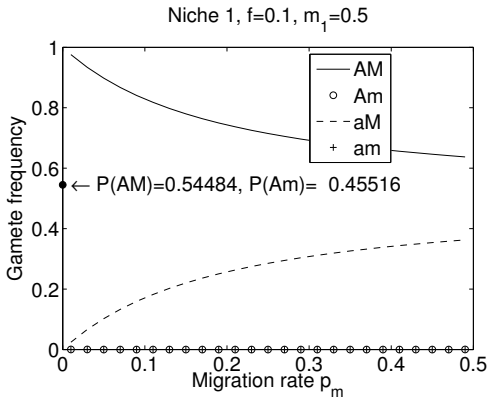



(b) Assortative mating ($n_0 = 0$, $n_1 = 0.5$) Under assortative mating, the population needs a long time span to reach equilibrium. Therefore, the axis ranges up to 20000 generations which compresses the graph at low generation values in a way that the population seems not to start from a uniform distribution, but in fact it does.

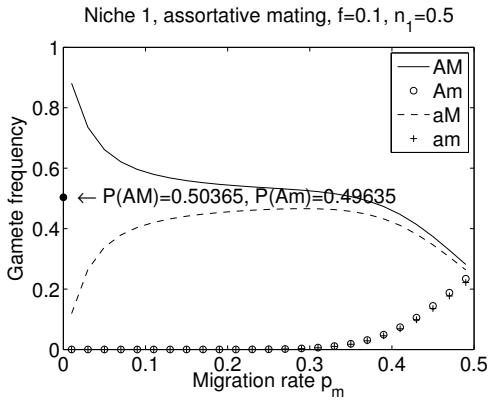



(c) Random mating

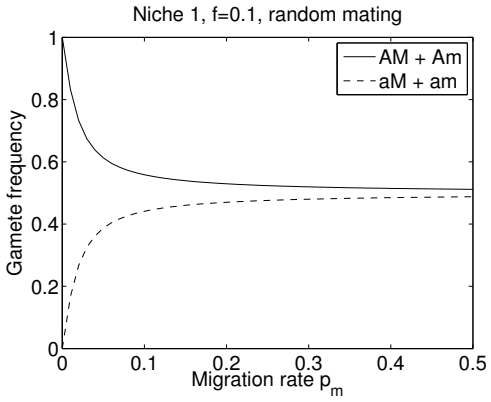
Figure 5.1: Iteration of the recurrence equations for uniform initial populations. The fitness difference is set to $f = 0.1$ and the migration rate is set to $p_m = 0.1$. Symbols are not uniform across the plots, but rather optimized to allow for better reading.




(a) Fitness-based mating ($m_0 = 0, m_1 = 0.5$)




(b) Assortative mating with ($n_0 = 0, n_1 = 0.5$)




(c) Random mating

Figure 5.2: The equilibrium values of gametes for various migration rates $p_m$. The fitness difference is set to $f = 0.1$ in all plots.

in populations under high migration pressures. In fact, we know from equation (4.165) that the profit of fitness-based maters is the higher, the stronger migration is.

Fitness-based mating permits for all $p_m$ the strongest 1st-locus polymorphism, and $M$ goes to fixation for all $p_m$.

For assortative mating, we notice that the difference between $AM$ and $aM$ is decreasing with higher $p_m$ and it does so, because $m$-carriers can co-exist with $M$-carriers. We know from equation (5.5) that assortative mating is not profitable for less fit individuals. Due to high migration, less fit individuals form a large part of the population and this large part sticks with random mating, i.e., the $m$-allele remains in the population.

For randomly mating populations, the migration rate has a minor influence for higher values ($p_m \geq 0.15$), but the population composition is more sensitive for changes of the migration rate below that threshold.

In the case without migration ($p_m = 0$), all three schemes allow a fixation of allele $A$ in niche 1 which is accomplished for roughly half of the population under fitness-based and assortative mating, or exactly half under random mating. Since there are no counteracting evolutionary forces, i.e., only selection is present, the equilibrium is reached quite fast for all mating schemes. From the time of $A$'s fixation, fitness-based mating loses its discriminating force and assortative mating looses its discriminating aim, because all $A$-carriers mate without any difference between random, assortative or fitness-based maters. Thus, also at the second locus, the selection pressure vanishes and alleles $M$ and $m$ co-exist for $p_m = 0$.

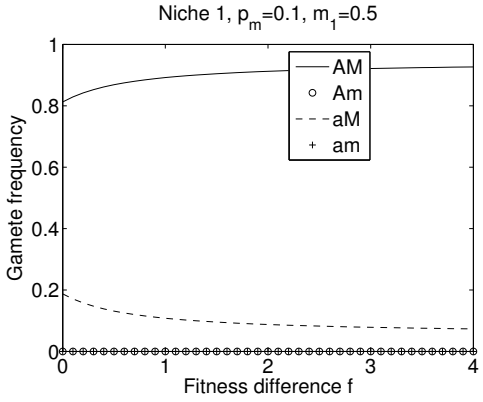
### 5.2.3 Ecological selection pressure

The strength of natural selection is quantified by the fitness advantage $f$ which $i^*$ has over the other allele. The stronger natural selection is, the higher the relative equilibrium frequency of $i^*$ is. Figure 5.3 confirms this and reveals that ecological selection has a marginal influence when $f$ is low for fitness-based mating, it has a rather strong influence for assortative mating, and it has a moderate influence for random mating.

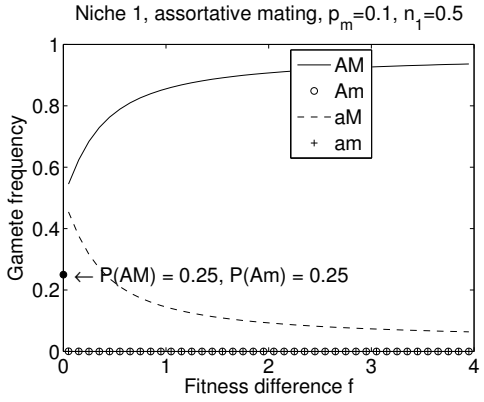
While fitness-based mating causes even for low fitness values a strong polymorphism, assortative mating results in similar levels of polymorphism only if natural selection is stronger than a certain threshold. For $f \geq 1$, the effect of the mating scheme (fitness-based or assortative) is negligible. Even random mating maintains comparable high levels of $i^*$ for that $f$-range, so that all mating schemes perform equally well.

### 5.2.4 Selection-Migration Equilibrium

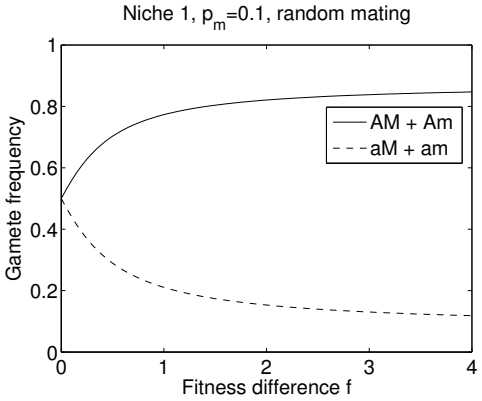
The equilibrium frequency of $i^*$, namely, $\hat{P}(i^*)$, is shown in figure 5.4 as a function of $p_m$ and $f$. With the exception of $p_m = 0$ or $f = 0$, all three mating schemes perform similar: Migration rates have higher influence than the fitness difference—at least for the $f$-range tested. But if $f = 0$ assortative and random mating show no difference in $\hat{P}(i^*)$, allele $i^*$ gets




(a) Fitness-based mating ($m_0 = 0, m_1 = 0.5$)

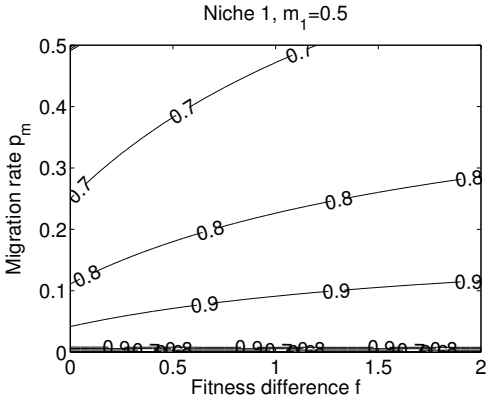



(b) Assortative mating ($n_0 = 0, n_1 = 0.5$)

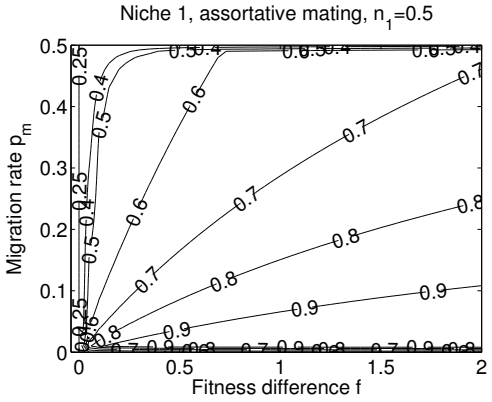



(c) Random mating

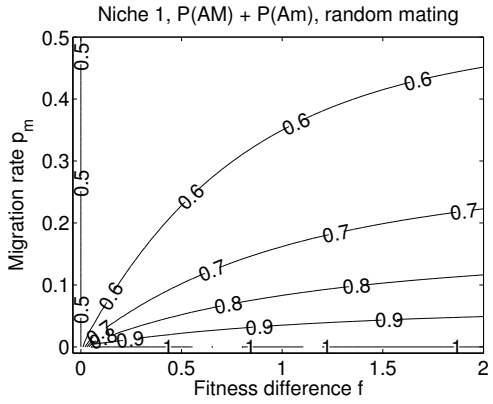
Figure 5.3: The relative frequency of gametes in the equilibrium for various fitness differences $f$ (x-axis). The migration rate is set to $p_m = 0.1$.




(a) Fitness-based mating ($m_0 = 0, m_1 = 0.5$)




(b) Assortative mating ($n_0 = 0, n_1 = 0.5$)




(c) Random mating

Figure 5.4: The relative frequency of gamete $AM$ in the equilibrium for various migration rates (y-axis) and fitness differences (x-axis). The labels of the contour lines show the value of $\hat{P}(AM)$.

half of the population irrespective of the migration rate. Under fitness-based mating, however, $i^*$ is maintained at fairly high levels even for high migration rates. This level decreases with $p_m$ for all mating schemes.

If $p_m = 0$, then the allele $A$ goes to fixation, and the allele $M$ and $m$ co-exist in all mating schemes. If $p_m > 0$, then the allele $M$ goes to fixation and mate preferences are exploited by the whole population.

The figure 5.4a shows that $P(AM)$ drops sharply for $f = 0$ and $p_m = 0.5$ as is seen at the contour lines in the upper left corner of the plot. In fact, for this specific parameter combination, the allele $M$ does not go to fixation. We can calculate the equilibrium frequencies for a homogeneous starting population. In this case, the frequencies after migration are given by

$$P^{\text{niche }1}(j,k) = \frac{1}{2}\Big(p^{\text{niche }1}(j,k) + p^{\text{niche }2}(j,k)\Big) = P^{\text{niche }2}(j,k) = \frac{1}{4}. \tag{5.6}$$

Using the recurrence equation (4.136), the mating probabilities in table 4.9, and equation (5.6), the relative gametic frequencies in the second generation are given by

$$p^{\text{niche }1}(AM) = p^{\text{niche }2}(aM) = \frac{1}{4} + m_1\frac{3}{32} \tag{5.7}$$

$$p^{\text{niche }1}(Am) = p^{\text{niche }2}(am) = \frac{1}{4} + m_1\frac{1}{32} \tag{5.8}$$

$$p^{\text{niche }1}(aM) = p^{\text{niche }2}(AM) = \frac{1}{4} - m_1\frac{3}{32} \tag{5.9}$$

$$p^{\text{niche }1}(am) = p^{\text{niche }2}(Am) = \frac{1}{4} - m_1\frac{1}{32}. \tag{5.10}$$

We see that all gametes will occur again at the same relative frequency of $\frac{1}{4}$ after migration. So, the equilibrium is attained already after two generations. Clearly, in this special case, $M$ does not go to fixation and we notice in figure 5.4a the sharp decline of $\hat{P}(AM)$ from roughly 1 to $\frac{1}{4} + m_1\frac{3}{32}$.

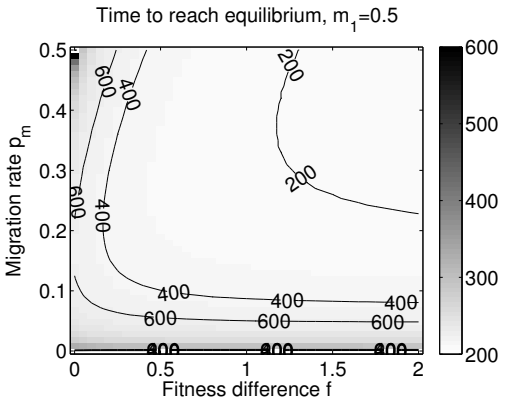
### 5.2.5 The time of approaching the equilibrium

Let $t_\epsilon$ be the earliest time at which the population distribution will differ only $\epsilon$ from those of the subsequent generation. I have predicted in section 5.1.2 that

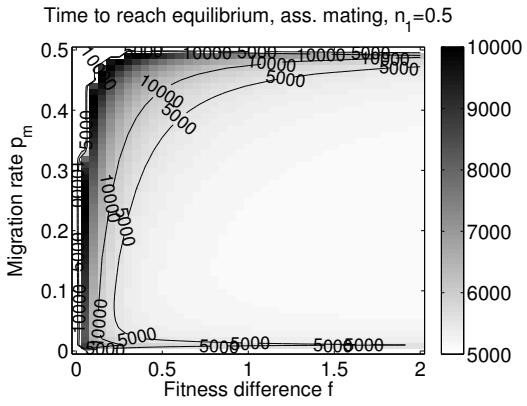
$$\begin{aligned} \frac{\partial t_\epsilon}{\partial p_m} &< 0 \\ \text{with } t_\epsilon &= \min_{t\in\mathbb{N}_0}\Big\{t:\ \sum_{(j,k)} |p_t^n(j,k) - p_{t+1}^n(j,k)| < \epsilon,\ \forall(j,k),\ \forall n\Big\} \end{aligned} \tag{5.11}$$

for a given $\epsilon$.

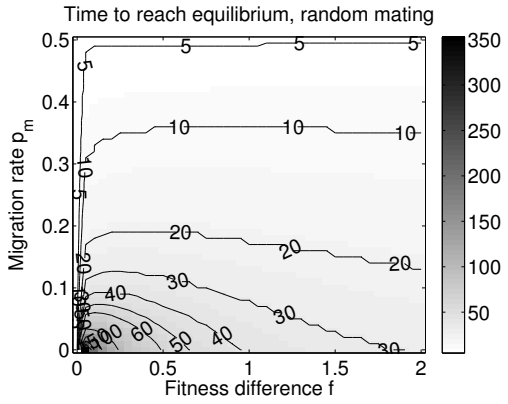
Figure 5.5 plots $t_\epsilon$ with $\epsilon = 10^{-6}$. The visualization challenge here is to collocate extremely inhomogeneous data. That is why I have chosen




(a) Fitness-based mating ($m_0 = 0, m_1 = 0.5$)

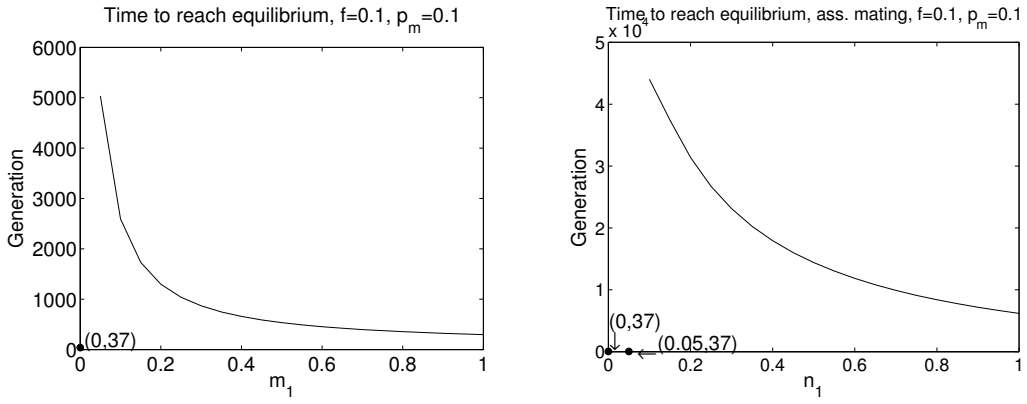



(b) Assortative mating ($n_0 = 0, n_1 = 0.5$)




(c) Random mating

Figure 5.5: The time $t_\epsilon$ until the equilibrium is reached for various migration rates (y-axis) and fitness differences (x-axis). The value of $t_\epsilon$ is shown in a combination of a contour plot and a grey-scaled image. The darker a region is, the higher $t_\epsilon$ is (see color coding to the right of the plots). The labels at the contour lines also show $t_\epsilon$.




(a) Fitness-based mating with $m_0 = 0$

(b) Assortative mating with $n_0 = 0$

Figure 5.6: The time $t_\epsilon$ until the equilibrium is reached for various mating tendencies $m_1/n_1$ (x-axis).

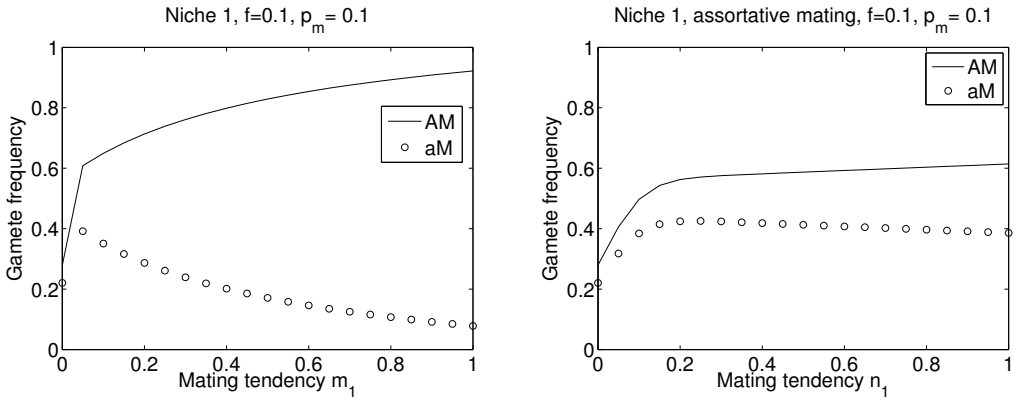
to combine two plot types. The matrix of $t_\epsilon$-values (given $p_m$ and $f$) is displayed as a grey-scaled image and with contour lines. In this way, abrupt $t_\epsilon$-changes are visible as well as the interrelation between $p_m$ and $f$ which defines the borderlines of regions with similar $t_\epsilon$.

The above-mentioned prediction, even though it was stated for sexual selection under fitness-based mating, is confirmed for both fitness-based and random mating. The picture for assortative mating (figure 5.5b) is somewhat less clear. There, we see that $t_\epsilon$ is firstly decreasing with $p_m$ and then it increases again.

For fitness-based mating, $t_\epsilon$ shows sharp changes in two cases. Firstly, when $f = 0$ and $p_m$ increasing from a value below 0.5 to the value 0.5, and secondly, for all $f$ when $p_m$ increases from zero to a positive value. The first case has already been treated in section 5.2.4. Homogeneous initial populations in combination with the parameters $f = 0$ and $p_m = 0.5$ cause a fast evolutionary stop. Consequently, also $t_\epsilon$ is rather low. In the second case, when $p_m = 0$, the allele $A$ goes to fixation while allele $M$ does not (this is also the case for assortative mating). However, when $p_m > 0$, then a constant influx of $a$-alleles maintains the advantage that fitness-based maters have over random maters. This leads to a fixation of $M$. The transition from the co-existence of $M$ and $m$ for $p_m = 0$ to the fixation of $M$ is reflected by a sharp change of $t_\epsilon$ which can be seen in an accumulation of contour lines at $p_m \approx 0$.

For random mating, the equilibrium is attained with increasing speed for increasing $f$ and $p_m$.

Figure 5.6 shows $t_\epsilon$ for different sexual selection strength (fitness-based, figure 5.6a, and assortative, figure 5.6b). The measure $t_\epsilon$ for assortative and fitness-based mating coincides with that for random mating when $m_1 = n_1 = 0$. Furthermore, we see at figure 5.6 that—except when $m_1 = 0$




(a) Fitness-based mating with $m_0 = 0$

(b) Assortative mating with $n_0 = 0$

Figure 5.7: The relative frequency of $M$-carriers ($m$-carriers become extinct) in the equilibrium for various mating tendencies $m_k$ (fitness-based mating, left plot) and $n_k$ (assortative mating, right plot). The x-axes give the $m_k/n_k$-values. The fitness advantage is set to $f = 0.1$ and the migration rate is set to $p_m = 0.1$.

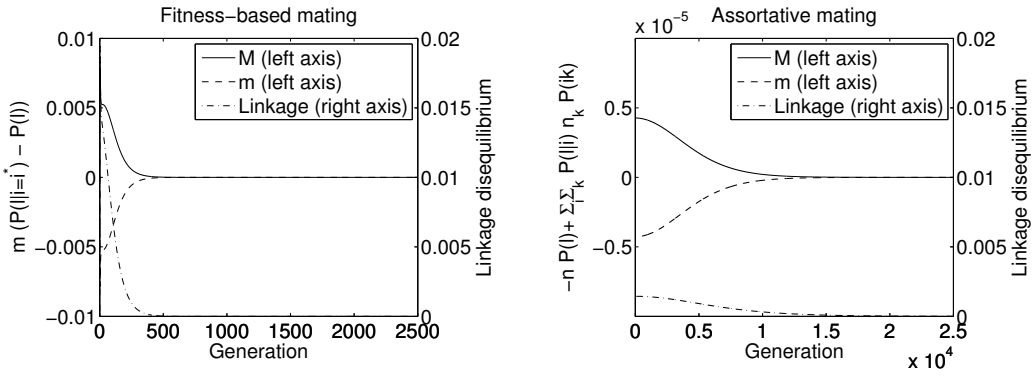
and $n_1 = 0$—the equilibrium is attained faster for stronger mate preferences, i.e., bigger $m_1/n_1$. For fitness-based mating, the equilibrium is reached at a roughly $10^4$ times higher speed than for assortative mating.

The low value of $t_\epsilon$ for $n_1 = 0.05$ under assortative mating is an artifact caused by an extremely low evolution. Even though the sum of differences between relative gametic frequencies in two subsequent generations differ less than $\epsilon = 10^{-6}$, the evolution does proceed. Letting $\epsilon$ decrease to $10^{-7}$ gives $t_\epsilon = 186815$ which resembles closer the actual approach of the equilibrium.

### 5.2.6 The strength of sexual selection

Fitness-based mating enhances natural selection. As a consequence, the relative equilibrium frequencies of $i^*$ should increase with $m_1$. Figure 5.7 plots $\hat{P}(i^*)$ for both mate preferences (fitness-based, figure 5.7a, and assortative, figure 5.7b) and confirms this. The relative equilibrium frequency $\hat{P}(i^*)$ increases with the sexual selection strength which is given by $m_1$ (or by $n_1$ in assortatively mating populations).

Even though both mate preferences go to fixation, fitness-based mating provides far higher levels of allele $i^*$ in the respective niches, thus, it maintains a stronger polymorphism. This was tested for $f = 0.1$ and $p_m = 0.1$.




(a) Fitness-based mating ($m_0 = 0, m_1 = \frac{1}{2}$)

(b) Assortative mating ($n_0 = 0, m_1 = \frac{1}{2}$)

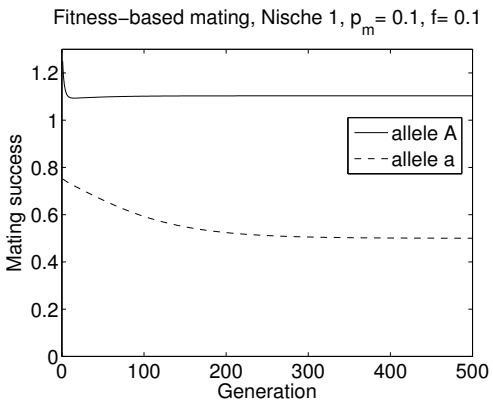
Figure 5.8: The difference of $P_{\text{chosen}}(l)$ to $P_t(l)$ and the linkage disequilibrium $D$ for fitness-based mating (left plot) and assortative mating (right plot) over the course of time (x-axis). The y-axes in each plot show the values of $P_{\text{chosen}}(l) - P_t(l)$ at the left axes, and the values of $D$ at the right axes. The fitness advantage is set to $f = 0.1$ and the migration is set to $p_m = 0.1$.

### 5.2.7 Mating probabilities for the second locus

I have derived in section 4.2.2 the probability for a 2nd-locus genotype $l$ of being chosen, see equation (4.169). The analogous probability for assortative mating is calculated using the equations (4.166), (4.176), and summing over $i$. We obtain

$$P^{\text{am}}_{\text{chosen}}(l) = P(l)(1 - \bar{n}) + \sum_{i,k} P(l|i) n_k P(i,k). \tag{5.12}$$

We see at both formulas, (4.169) and (5.12), that $P_{\text{chosen}}(l)$ can differ from $P_t(l)$. This is crucial, because it means that mating is not random at the second locus. Figure 5.8 plots the difference, i.e., $P_{\text{chosen}}(l) - P_t(l)$, for both alleles $M$ and $m$ under fitness-based and under assortative mating. We see that there is a linkage disequilibrium in the beginning of the evolution, and this causes $P_{\text{chosen}}(l)$ to be different from $P_t(l)$. In both mating schemes, the allele $M$ is more often chosen than its relative frequency would predict. The linkage disequilibrium $D$ is around five times higher for fitness-based mating than for assortative mating and so is the difference to $P_t(l)$. This means that allele $M$ co-occurs in the fitness-based mating model more often with $i^*$. This increases $P_{\text{chosen}}(M)$ and assists the fixation of $M$. As a consequence, fitness-based mating populations reach the equilibrium that fast. Both mating schemes lead to a linkage-free population in equilibrium.




(a) Fitness-based mating ($m_0 = 0, m_1 = \frac{1}{2}$)

Figure 5.9: The relative mating success [29] is plotted in each time step for males under fitness-based mating ($m_0 = 0, m_1 = \frac{1}{2}$) as given in equation (4.175).

### 5.2.8 Gavrilets' measure of mating success

Another measure for the effect of non-random mating is the relative mating success as defined by Gavrilets [29], see section 4.2.3. I have calculated that measure for fitness-based mating in equation (4.175). For assortative mating, the corresponding values for females and males are given by

$$w_{f,jk} = 1, \quad w_{m,il} = 1-\bar{n}+\sum_k n_k P(k|i) = 1+\frac{1}{P(i)}\sum_k n_k\Big(P(i,k)-P(i)P(k)\Big). \tag{5.13}$$

If the linkage disequilibrium $D$ is equal to zero, then $w_{m,il} = 1$ holds. We have seen for assortative mating in the previous section 5.2.7 that $D$ is marginally different from zero and converges to zero, and hence, the relative mating success for males is marginally different from 1 under assortative mating. Therefore, I only plot the male relative mating success for fitness-based mating in figure 5.9. We see that allele $A$ has generally higher relative mating success than allele $a$. The relative mating success converges to $1 - m_1$ for $i \neq i^*$ and $1 + m_1(1 - \hat{P}(i^*))/\hat{P}(i^*)$ for $i^*$-males as the population approaches its equilibrium, see equation (4.175).

## 5.3 Simulations with finite population sizes

The qualitative behavior of a model depends on whether the population size is assumed to be infinite or finite. In this and the following sections, fitness-based mating is implemented for a finite population by using agent-based modeling.

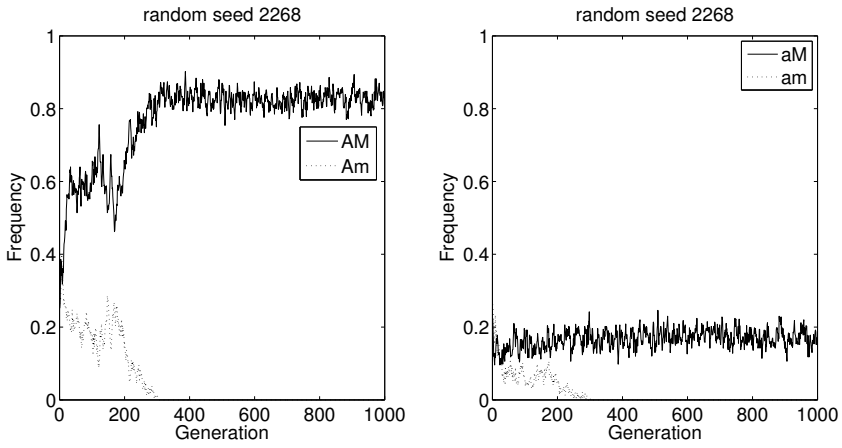



Figure 5.10: Typical run of an agent-based simulation of the fitness-based mating model. Parameters: $N_1 = N_2 = 500$, 5000 generations (only the first 1000 are shown), $p_m = 0.1$, $f = 0.1$, $m_0 = 0$, $m_1 = 0.5$. The left plot shows the frequencies of gametes $AM$ and $Am$, the right plot those of gametes $aM$ and $am$.

Three agent-based models are examined in this section. The first of the two finite-size implementation is a straightforward realization of the fitness-based mating model (section 5.3.2). Each element is an individual agent whose fitness and mating behavior is determined by its genes. The second model allows the evolution of the values of $m_k$ (section 5.3.3). The third finite-size implementation is an analogous version of the model of Dieckmann and Doebeli [16] (section 5.3.4). The difference to the model of Dieckmann and Doebeli is that the mating character now codes the tendency for fitness-based mating instead of assortative mating.

### 5.3.1 Notes on the computer implementation

Software implementations generally use pseudo-random numbers which influence the reproducibility of simulations. In the following implementations, the random number generator provided by *JAVA*™ (class `Random`) which uses a linear congruential pseudo-random number generator has been used.

The models have been implemented in the *JAVA*™ programming language and compiled with the `javac`-compiler of version *1.5.0_06*. The programs were run on a *Sun Fire 880*-machine with a *Sparc US-III*, 64-bit processor and the *SunOS 5.10* operation system.

### 5.3.2 Finite-size implementation of the fitness-based mating model

The model consists of two populations, one population for each niche. These populations are initialized randomly. That means, each element is placed in one of the two niches with equal likelihood and each allele of

both haploid genes is chosen from a uniform distribution. Therefore, the initial genotype distributions are binomial.

The reproduction procedure is implemented differently compared to the case that assumes infinite population sizes. In this model version with finite population sizes, the number of offspring for each couple is the sum of the fitness values of the parents. After all elements have reproduced, selection eliminates randomly chosen offspring as long as the capacity of the niche is exceeded. This implementation approximates the offspring production in the case of infinite population sizes.

Figure 5.10 presents the result of a typical simulation. It shows that the qualitative behavior does not differ significantly from the deterministic case, see figure 5.1a, even though finite-size effects are present and also visible in the fluctuations. The finite-size effects are the more pronounced, when the niche sizes are smaller. However, it seems that fitness-based mating somewhat lowers stochastic uncertainties, because even simulations with small niche sizes ($N_1 = N_2 = 50$) maintain a polymorphism over a long period. Stochastic fluctuations lead sooner or later to the fixation of either allele $A$ or allele $a$. However, running a simulation for 5,000,000 generations still shows a polymorphism. This suggests that fitness-based mating makes the dynamics more robust against fluctuations, presumably by its effect of enhancing natural selection.

### 5.3.3 Evolution of fitness-based mating

The finite-size model is changed in a way that each individual now has its own individual value of $m_1$. The value for $m_0$ is fixed to 0, since fitness-based mating shall compete with random mating, which would not be the case for positive $m_0$. Each parent passes its $m_1$-value with a probability of 50% to an offspring where it is mutated with probability $p_{m_1}$. In the case of a mutation, $m_1$ is either increased or decreased by $p_{\text{inc}}$: $m_1 \pm p_{\text{inc}}$ with equal probabilities.

An open question is which levels of $m_1$ would result when also the $m_1$-value is under evolution. According to the analytical result, see equation (4.141), individuals with higher $m_1$ have higher reproductive success as long as $i^*$ is not fixed. Since there is a polymorphism in each niche, $i^*$ will not fix, thus, there is an enduring advantage for higher $m_1$-individuals. Therefore, the mean value of $m_1$ in the population should increase until a mutation-selection equilibrium is reached.

The mutation-selection equilibrium of $\bar{m}_1$ depends on $p_{m_k}$ and $p_{\text{inc}}$ as can be seen in the plots of figure 5.11. There, a typical simulation run is shown, where the relative gamete frequencies in niche 1 are shown on the left side and the population mean of $m_1$ on the right side. Upper plots are run with parameters $p_{m_1} = 0.05$ and $p_{\text{inc}} = 0.1$, whereas in the lower plots the probability $p_{\text{inc}}$ is changed to $p_{\text{inc}} = 0.01$.

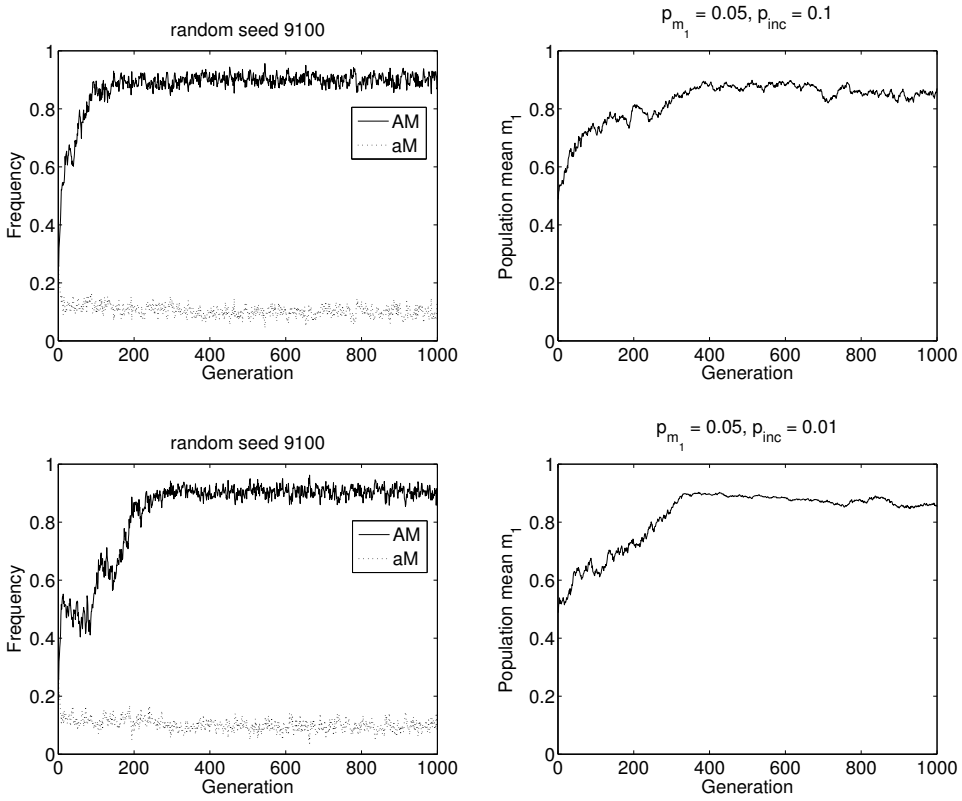



Figure 5.11: Typical run of an agent-based simulation of the fitness-based mating model. Parameters: $N_1 = N_2 = 500$, 5000 generations (only the first 1000 are shown), $p_m = 0.1$, $f = 0.1$, $m_0 = 0$, $m_1 = 0.5$. The top plots show a simulation with $p_{m_1} = 0.05$ and $p_{\text{inc}} = 0.1$, lower plots show a simulation with $p_{m_1} = 0.05$ and $p_{\text{inc}} = 0.01$. On the left, relative frequencies of gametes $AM$ and $aM$ are shown (others become extinct and are not displayed for the sake of better readability). On the right, the population mean of $m_1$ is plotted.

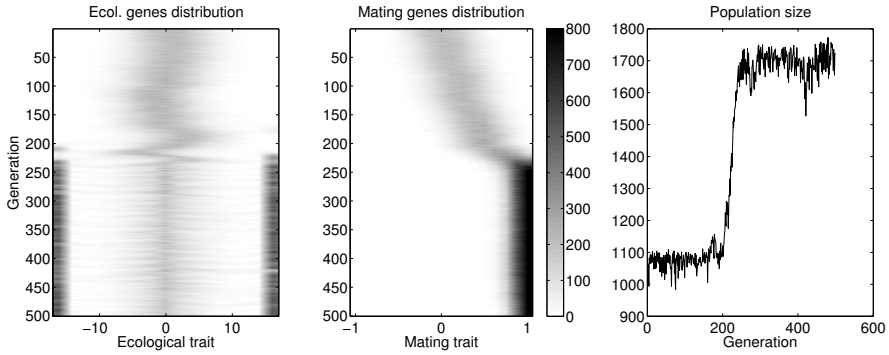



Figure 5.12: Typical run of a stochastic implementation of the Dieckmann-Doebeli model with modified mate preferences. The mating genes code for the tendency to mate fitness-based instead of for assortative mating. Parameters: Initial population size 1000, 500 generations, $p_m = 0.001$, $\sigma_K = 1$, $\sigma_C = 0.15$, $r$=1, $K_0 = 1000$, mating traits $= (-1, -\frac{7}{8}, \ldots, 0, \ldots, \frac{7}{8}, 1)$. Genes of initial population members are initialized such that each allele is drawn from a uniform distribution.

### 5.3.4 Modified Dieckmann-Doebeli model

Section 3.4.1 introduced the speciation model of Dieckmann and Doebeli [16]. In this subsection, the simulation of a modified version of the Dieckmann-Doebeli model is presented. The modifications regard the mating behavior and the reproduction procedure.

An individual consists of two sets of eight diploid genes. The first set determines the ecological trait and the second set determines the mating behavior. Ecological genes take alleles $A$ or $a$ and the ecological trait is calculated by the difference between the number of $A$- and $a$-alleles. Therefore, it ranges between $-16$ to 16 in steps of 2. Instead of assortative or non-assortative mating, the set of mating genes determines now fitness-based or non-fitness-based mating. The mating trait $k$ is calculated in the same way as the ecological trait and also ranges between $k = -16$ to $k = 16$ in steps of 2. Each $k$ is associated with a tendency $m_k$, which ranges from $-1$ to 1 in steps of 1/8. Negative $m_k$ correspond to "anti-fitness-based mating" which is a preference for the least fit individuals. If $m_k = -1$, then the individual mates with the least fit mate available. If $m_k = -1/8$, then the individual mates in 87.5% of the cases randomly and in 12.5% of the cases with the least fit available. If $m_k = 0$, then the individual mates randomly. A value $m_k > 0$ corresponds to the probability to mate fitness-based.

Further adjustments are made in order to achieve a population branching. The population is monogamous. Elements are randomly assigned to each gender and are ordered according to their fitness value in two lists, one list for females, one for males. Fittest individuals appear first in the list. The mating procedure is such that the fittest female chooses among

the males, both reproduce, and leave the lists, i.e., the mating pool. Then the next fittest female gets its turn. If she mates anti-fitness-based, then she chooses the last element in the list. If she mates randomly, then she chooses a male at a random position in the list.

The number of offspring is determined as in the Dieckmann-Doebeli model, where resource availability, population composition, and competition between population members influence the growth rate of a genotype. Let $N(j,k)$ be the number of population elements with genotype $(j,k)$ at a certain time. Its growth rate to the subsequent generation is

$$\frac{\partial N(j,k)}{\partial t} = rN(j,k)\left(1 - \frac{\sum_{(i,l)} N(i,l)C(i,j)}{K(j)}\right), \tag{5.14}$$

where

$$\begin{aligned} r &\quad \text{is the birth rate} \\ K(j) &= K_0 \exp\left(-\frac{j^2}{2\sigma_K^2}\right) \text{ are the resources available for } j \\ C(i,j) &= \exp\left(-\frac{(i-j)^2}{2\sigma_C^2}\right) \text{ is the competition between } i \text{ and } j. \end{aligned}$$

There are four differences to the original fitness-based model:

- Instead of two niches, this model has only one niche.
- All individuals are monogamous, instead of monoandrous females and polygynous males.
- Non-fitness-based mating is incorporated.
- A mating sequence is introduced, which privileges fit elements to choose first.

Figure 5.12 presents a typical simulation run. The left plot gives the evolution of the ecological trait. The $y$-axis plots the time and the $x$-axis the value of the ecological trait. The darker a point in the plot is, the more population members carry that trait. The evolution of the second trait is shown in the middle plot. The population moves from intermediate mating traits (random mating) to positive mating traits (fitness-based mating). Once a significant part of the population mates fitness-based, the population splits into two branches. One branch consists of elements with negative ecological traits and one branch of positive ecological traits. A minority of the population colonizes intermediate ecological trait values.

The splitting into the two major clusters is caused by the fact that matings take place between elements of the same fitness value, that means fit elements pair among themselves and less fit elements also pair among themselves. Less fit elements do not find any fit mates left in the mating pool. This separation of the mating pool is caused by the mating procedure, in particular by the mating sequence and monogamy. Each fitness

group mates only within its own fitness class. Both group sizes are stabilized by negative frequency-dependent selection such that the smaller group experiences less competition and can multiply in numbers, whereas the bigger group experiences more competition and does not reproduce that often.

Equal fitness values imply equal phenotypes thus, assortative mating is enforced. In such way the population splits, because assortative mating leads to an accumulation of homozygotic genes. This model is an example that fitness-based mating can coincide with assortative mating, even when there is genetic variance present in the population.

The existence of separate mating pools corresponds to the reproductive isolation between those pools, which however, is not complete. Since the population branches represent genetic and ecological clusters which form moreover a reproductive community, the population branches correspond to incipient species.

The selection pressure for fitness-based mating is presumably maintained and the branched population is stable. Simulations over 50,000 generations support this statement.

# 6 Summary

**Introduction of fitness-based mating.** I have proposed a mate preference and I have coined it *fitness-based mating* (chapter 4). It is inspired by nature, where individuals apparently include physical, behavioral, or mental properties of potential mates in their mating decisions (section 2.4.2).

Fitness-based mating has been modelled in population-genetic terms which allows analytical treatment and considerable statements (section 4.2). The analytical predictions (sections 4.1.7 and 4.2.2) have been computationally verified (section 5.2).

Fitness-based mating is an evolutionary successful mating strategy. It spreads in a population due to its amplifying effect on the reproductive success and on the attractiveness of its carriers (section 4.2.2). The explicit equilibrium distribution has been derived (section 4.1.7).

**Fitness-based mating vs. assortative mating.** Fitness-based mating has been compared to assortative mating (biologically: section 2.4, analytically: section 4.2.4), because assortative mating is believed to play a role in speciation processes. From an evolutionary perspective, fitness-based mating is not only more plausible, because it is more beneficial in terms of reproductive success, but it also appears to be an easier strategy to use. Similarity evaluations for assortative mating require some degree of self-knowledge whereas fitness evaluations can be embedded in mating rituals. The latter seems to be realized in several biological species (section 2.4.2).

Fitness-based mating is an alternative to assortative mating, but it can coincide with assortative mating, for instance, when all niche elements have the fittest allele (section 4.2.4) or when fitness-based mating is implemented in a specific way (section 5.3.4).

**Fitness-based mating without ecological selection.** The effect of fitness-based mating has been investigated when natural selection is removed (analytically: section 4.1.7, setting $f = 0$ in equation (4.121), computationally: section 5.2.4, again setting $f = 0$). The removal of ecological selection allows the identification of specific properties of sexual selection induced by mate preferences. In contrast to assortative mating, fitness-based mating maintains the polymorphism at a higher level even in the absence of ecological selection.

**Interplay between natural and sexual selection.** The combined effects of fitness-based mating and natural selection have been examined (mathematically: section 4.2.2, numerically: section 5.2). On the individual level,

fitness-based mating increases the reproductive success and the attractiveness of its carrier and therefore, it is applied by an increasing number of population members. On the population level, fitness-based mating allows higher fractions of well adapted individuals than under random or assortative mating (section 5.2.4 and 5.2.6); and fitness-based mating provides faster adaptation to ecological conditions than under assortative mating (sections 5.2.5).

Computer simulations suggest that fitness-based mating induces robustness. It reduces the impact of stochastic fluctuations, in terms of reducing extinction probabilities (section 5.3.2). Thus, sexual selection induced by fitness-based mating increases the survival chances of the whole population.

**Fitness-based mating and speciation.** Fitness-based mating might also play a role in speciation processes, since a polymorphism arises naturally in the model of fitness-based mating. A polymorphism is the result of a balance of evolutionary factors such as natural selection, sexual selection, and migration. Migration counteracts both selection schemes in the fitness-based mating model. I quantify the polymorphism of two alleles by the difference between the relative frequencies of the fittest allele and the least fit allele. Fitness-based mating causes a more pronounced polymorphism than random mating and assortative mating (sections 5.2.4 and 5.2.6).

The emergence of a stable polymorphism of traits underlying ecological selection is of special interest, because a polymorphism can be a precursor of speciation. A polymorphic population corresponds to a clustered population in the genotype and phenotype space. Since fitness-based mating amplifies and stabilizes genotype and phenotype clusters by maintaining a strong polymorphism, fitness-based mating paves the way for speciation.

The ingredient missing for speciation is reproductive isolation between the clusters. When a population is clustered but its elements still interbreed, then the development of reproductive isolation could be triggered. Once reproductive isolation is present, speciation is completed. Fitness-based mating supports the clustering of the population, but fitness-based mating does not strengthen the reproductive isolation between subpopulations. So, if fitness-based mating acts together with reproductive-isolation inducing mechanisms, it can be part of speciation processes.

**Summary.** In this thesis, I have proposed and analyzed a population-genetic model that leads to a stable polymorphism through fitness-based mating. The model allows analytical treatment. The stability of the polymorphism is caused by the sexual selection through fitness-based mating, which both enhance the natural selection pressure. Due to the pronounced polymorphism caused by fitness-based mating, the fitness-based mate preference can be a part of speciation processes.

# A Notation

**Model parameters**

| | |
|---|---|
| $n$ or $\bar{n}$ | niche indices |
| $x_i$ or $x_i^{\text{niche n}}$ | a gamete, if needed with niche index. Equivalently used as the frequency of that gamete in the population |
| $i$, $j$, or $o$ | 1st-locus genotypes, counts the number of $A$-alleles |
| $l$, $k$, or $p$ | 2nd-locus genotypes, counts the number of $M$-alleles |
| $i^*$ or $j^*$ | fittest 1st-locus genotype in specific niche, $j^*$ denotes a different genotype in different niches |
| $m_k$ | probability to apply fitness-based mating, when individual has 2nd-locus genotype $k$ |
| $n_k$ | probability to apply assortative mating, when individual has 2nd-locus genotype $k$ |
| $p_m$ | migration rate between two niches |

**Fitness values and reproductive success**

| | |
|---|---|
| $f(j,k)$ | fitness value of genotype $(j,k)$ |
| $\phi(j,k)$ | reproductive success of genotype $(j,k)$ in current generation |
| $F(i,j)$ | couple fitness of parents with 1st-locus genotype $i$ and $j$ |
| $w_{f,jk}$, $w_{m,jk}$ | female and male mating success of $(j,k)$ as defined by [29] |

**Frequencies**

| | |
|---|---|
| $p(j,k)$ | frequency of gamete $(j,k)$ in the population prior to migration |
| $P(j,k)$ | frequency of gamete $(j,k)$ in the population after to migration |

**Probabilities**

| | |
|---|---|
| $P_{\text{meet}}(il,jk)$ | meeting probability of $(i,l)$ and $(j,k)$ |
| $P_{\text{mate}}(il,jk)$ | mating probability of $(i,l)$ and $(j,k)$ |
| $P_{jk}^{\text{choice}}(il)$ | probability that $(j,k)$ chooses $(i,l)$ for mating |
| $P_{il}^{\text{accept}}(jk)$ | probability that $(i,l)$ accepts a mating with $(j,k)$ |
| $P_{\text{off}}(op\|il,jk)$ | probability that an offspring has the genotype $(o,p)$ when its parents are $(j,k)$ and $(i,l)$ |
| $P_{\text{chosen}}(i,l)$ | probability of $(i,l)$ to be chosen for mating |

**Other**

| | |
|---|---|
| $D$ | linkage disequilibrium |
| $\mathfrak{A}$ | set of all possible genotypes |
| $t_\epsilon$ | time, at which the equilibrium is reached given a predefined maximal distance $\epsilon$ to the actual equilibrium distribution |

# Bibliography


[1] R. C. Albertson, J. T. Streelman, and T. D. Kocher. Directional selection has shaped the oral jaws of lake malawi cichlid fishes. *Proceedings of the National Academy of Sciences of the United States of America*, 100(9):5252–5257, 2003.

[2] J. Alcock. *Animal Behavior: An evolutionary approach*. Sinauer Associates, 1998.

[3] P. Arcese and J. N. M. Smith. Effects of population density and supplemental food on reproduction in song sparrows. *Journal of Animal Ecology*, 57(1):119–136, 1988.

[4] A. J. Batesman. Intra-sexual selection in *drosophila*. *Heredity*, 2:349–368, 1948.

[5] C. T. Bergstrom and L. A. Real. Towards a theory of mutual mate choice: Lessons from two-sided matching. *Evolutionary Ecology Research*, 2:493–508, 2000.

[6] O. Breidbach and J. Jost. Working in a multitude of trends – Species balancing populations. *Journal of Zoological Systematics and Evolutionary Research*, 42(3):202–207, 2004.

[7] E. D. Brodie III. Correlational selection for color pattern and antipredator behavior in the garter snake *thamnophis ordinoides*. *Evolution*, 46(5):1284–1298, 1992.

[8] W. D. Brown. The cause of size-assortative mating in the leaf beetle *trirhabda canadensis* (coleoptera: Chrysomelidae). *Behavioral Ecology and Sociobiology*, 33:151–157, 1993.

[9] N. A. Campbell. *Biologie*. Spektrum, 1997.

[10] T. H. Clutton-Brock, S. D. Albon, R. M. Gibson, and F. E. Guinness. The logical stag: Adaptive aspects of fighting in red deer (*cervus elaphus l.*). *Animal Behaviour*, 27(1):211–225, 1979.

[11] A. Cockburn. *Evolutionsökologie*. Gustav Fischer Verlag, 1995.

[12] D. Cook. Sexual selection in dung beetles. 2. Female fecundity as an estimate of male reproductive success in relation to horn size, and alternative behavioral strategies in *onthophagus-binodis thunberg* (scarabaeidae, onthophagini). *Australian Journal of Zoology*, 36:521–532, 1988.

[13] C. Darwin. *The Origin of Species*. Penguin, 1876.

[14] C. Darwin. *The Descent of Man, and Selection in Relation to Sex*. Penguin books, 2004.

[15] H. Dickinson and J. Antonovics. Theoretical considerations of sympatric divergence. *The American Naturalist*, 107(954):256–274, 1973.

[16] U. Dieckmann and M. Doebeli. On the origin of species by sympatric speciation. *Nature*, 400(6742):354–357, 1999.

[17] A. F. Dixson, T. Bossi, and E. J. Wickings. Male dominance and genetically determined reproductive success in the mandrill (*mandrillus sphinx*). *Primates*, 34(4):525–532, 1993.

[18] M. Drés and J. Mallet. Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, 357(1420):471–492, 2002.

[19] L. A. Dugatkin. Sexual selection and imitation: Females copy the mate choice of others. *The American Naturalist*, 139(6):1384–1389, 1992.

[20] L. A. Dugatkin and J.-G. J. Godin. Reversal of female mate choice by copying in the guppy (*poecilia reticulata*). *Proceedings: Biological Sciences*, 249(1325):179–184, 1992.

[21] D. J. Emlen. Costs and the Diversification of Exaggerated Animal Structures. *Science*, 291(5508):1534–1536, 2001.

[22] J. A. Endler. *Geographic Variation, Speciation, and Clines*. Princeton University Press, 1977.

[23] I. Eshel. Evolutionary and continuous stability. *Journal of Theoretical Biology*, 103:99–111, 1983.

[24] W. J. Ewens. *Mathematical Population Genetics I. Theoretical Introduction*. Springer-Verlag, New York, 2004.

[25] A. Fabiani, F. Galimberti, S. Sanvito, and A. R. Hoelzel. Extreme polygyny among southern elephant seals on Sea Lion Island, Falkland Islands. *Behavioral Ecology*, 15(6):961–969, 2004.

[26] J. Felsenstein. Skepticism towards Santa Rosalia, or why are there so few kinds of animals. *Evolution*, 35(1):124–138, 1981.

[27] J. Fickel, D. Lieckfeldt, P. Ratanakorn, and C. Pitra. Distribution of haplotypes and microsatellite alleles among Asian elephants (*elephas maximus*) in Thailand. *European Journal of Wildlife Research*, 298-303(4):298–303, 2007.

[28] R. A. Fisher. *The Genetical Theory of Natural Selection*. Oxford University Press, 2000.

[29] S. Gavrilets. *Fitness Landscapes and the Origin of Species*. Princeton University Press, 2004.

[30] S. A. H. Geritz, E. Kisdi, G. Meszena, and J. A. J. Metz. Evolutionary singular strategies and the adaptive growth and branching of the evolutionary tree. *Evolutionary Ecology*, 12:35–57, 1998.

[31] S. F. Gilbert. Opening Darwin's black box: teaching evolution through developmental genetics. *Nature Reviews: Genetics*, 4:735–741, 2003.

[32] D. E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine learning*. Addison Wesley, Reading, Massachusetts, 1989.

[33] P. R. Grant and B. R. Grant. Predicting microevolutionary responses to directional selection on heritable variation. *Evolution*, 49(2):241–251, 1995.

[34] M. R. Gross. Disruptive selection for alternative life histories in salmon. *Nature*, 313:47–48, 1985.

[35] M. R. Gross. Alternative reproductive strategies and tactics: diversity within sexes. *Trends in Ecology and Evolution*, 11:92–97, 1996.

[36] G. H. Hardy. Mendelian proportions in a mixed population. *Science*, 28:49–50, 1908. Letter to the editor.

[37] H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, A. Hoang, C. E. Hill, P. Beerli, and J. G. Kingsolver. Strength and tempo of directional selection in the wild. *Proceedings of the National Academy of Sciences of the United States of America*, 98(16):9157–9160, 2001.

[38] J. Hofbauer and K. Sigmund. *Evolutionary games and replicator dynamincs*. Cambridge University Press, 1998.

[39] J. Höglund and A. Lundberg. Sexual selection in a monomorphic lek-breeding bird: Correlates of male mating success in the great snipe *gallinago media*. *Behavioral Ecology and Sociobiology*, 21:211–216, 1987.

[40] J. H. Holland. *Adaption in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial intelligence*. MIT Press, Cambridge, Massachusetts, 1975.

[41] A. Houtman and J. Falls. Negative assortative mating in the white-throated sparrow, *zonotrichia albicollis*: The role of mate choice and intra-sexual competition. *Animal Behaviour*, 48(2):377–383, 1994.

[42] S. K. Huber, L. F. De Lean, A. P. Hendy, E. Bermingham, and J. Podos. Reproductive isolation of sympatric morphs in a population of Darwin's finches. *Proceedings of the Royal Society B: Biological Sciences*, 274(1619):1709–1714, 2007.

[43] R. D. Hughes. Population dynamics of the cabbage aphid, brevicoryne brassicae (L.). *Journal of Animal Ecology*, 32(3):393–424, 1963.

[44] R. A. Johnstone. Sexual selection, honest advertisment and the handicap principle: Reviewing the evidence. *Biological Reviews*, 1995.

[45] K. N. Jones and J. S. Reithel. Pollinator-mediated selection on a flower color polymorphism in experimental populations of *antirrhinum* (Scrophulariaceae). *American Journal of Botany*, 88(3):447–545, 2001.

[46] J. Jost. On the notion of fitness, or: The selfish ancestor. *Theory in Biosciences*, 121:331–350, 2003.

[47] P. Kappeler. *Verhaltensbiologie*. Springer, 2006.

[48] P. Kappeler and C. v. Schaik. Sexual selection in Primates: Review and selective preview. In P. Kappeler and C. v. Schaik, editors, *Sexual selection in Primates - New and Comparative Perspectives*, pages 3–23. Cambridge University Press, 2004.

[49] E. L. Kessel. The mating activities of balloon flies. *Systematic Zoology*, 4(3):97–104, 1955.

[50] M. Kirkpatrick. Sexual selection and the evolution of female choice. *Evolution*, 36(1):1–12, 1982.

[51] R. Lande. Models of speciation by sexual selection on polygenic traits. *Proceedings of the National Academy of Sciences of the United States of America*, 78(6):3721–3725, 1981. Part 2: Biological Sciences.

[52] B. J. Le Boeuf. Male-male competition and reproductive success in elephant seals. *American Zoologist*, 14(1):163–176, 1974.

[53] H. Levene. Genetic equilibrium when more than one ecological niche is available. *The American Naturalist*, 87(836):331–333, 1953.

[54] R. C. Lewontin. *The Genetic Basis of Evolutionary Change*. Columbia University Press, 1974.

[55] D. Liebers, P. d. Knijff, and A. J. Helbig. The herring gull complex is not a ring species. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 271(1542):893–901, 2004.

[56] T. Malthus. *An Essay on the Principle of Population*. J. Johnson in St. Paul's Church-Yard, London, 1798. An Essay on the Principle of Population, as it Affects the Future Improvement of Society with Remarks on the Speculations of Mr. Godwin, M. Condorcet, and Other Writers.

[57] C. Matessi, A. Gimelfarb, and S. Gavrilets. Long-term buildup of reproductive isolation promoted by disruptive selection: How far does it go? *Selection*, 2(1-2):41–64, 2001.

[58] J. Maynard Smith. Sympatric speciation. *The American Naturalist*, 100(916):637–650, 1966.

[59] E. Mayr. The biological meaning of species. *Biological Journal of the Linnean Society*, 1:311–320, 1969.

[60] G. Mendel. Versuche über Pflanzenhybriden. In *Verhandlungen des naturforschenden Vereines in Brünn*, Bd. IV, 1865.

[61] T. C. M. Milinski, Manfred Bakker. Female sticklebacks use male coloration in mate choice and hence avoid parasitized males. *Nature*, 344,(6264):330–333, 1990.

[62] W. S. Moore. A single locus mass-action model of assortative mating, with comments on the process of speciation. *Heredity*, 42:173–186, 1979.

[63] T. Nagylaki. Continuous selective models with mutation and migration. *Theoretical Population Biology*, 5(2):284–295, 1974.

[64] T. Nagylaki. *Selection in one- and two-locus systems*. Lecture Notes in Biomathematics 15. Springer, 1977.

[65] T. Nagylaki and J. F. Crow. Continuous selective models. *Theoretical Population Biology*, 5(2):257–283, 1974.

[66] P. O'Donald. Assortive mating in a population in which two alleles are segregating. *Heredity*, 15:389–396, 1960.

[67] S. Pääbo, H. Poinar, D. Serre, V. Jaenicke-Després, J. Hebler, N. Rohland, M. Kuch, J. Krause, L. Vigilant, and M. Hofreiter. Genetic analysis from ancient DNA. *Annual Review Genetics*, 38:645–679, 2004.

[68] G. A. Parker. Mate quality and mating decision. In P. P. G. Bateson, editor, *Mate Choice*, pages 141–166, 1983.

[69] J. Pemberton, S. Albon, F. Guinness, T. H. Clutton-Brock, and G. Dover. Behavioral estimates of male mating success tested by DNA fingerprinting in a polygynous mammal. *Behavioral Ecology*, 3(1):66–75, 1992.

[70] P. S. Pennings, M. Kopp, G. Meszena, U. Dieckmann, and J. Hermisson. An analytical tractable model for competitive speciation. *American Naturalist*, 171:E44–E71, 2008.

[71] M. Petrie, T. Halliday, and C. Sanders. Peahens prefer peacocks with elaborate trains. *Animal Behaviour*, 41(2):323–331, 1991.

[72] R. A. Price and S. G. Vandenberg. Spouse similarity in American and Swedish couples. *Behavior Genetics*, 10(1):59–71, 1980.

[73] D. E. Promislow. Costs of sexual selection in natural populations of mammals. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 247(1320):203–210, 1992.

[74] D. E. Promislow, R. Montgomerie, and T. E. Martin. Mortality costs of sexual dimorphism in birds. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 250(1328):143–150, 1992.

[75] S. H. Rice. *Evolutionary Theory, Mathematical and Conceptual Foundations*. Sinauer Associates, 2004.

[76] N. A. Rosenberg and M. Nordborg. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews: Genetics*, 3:380–390, 2002.

[77] J. Ryan, Michael. Sexual selection, sensory systems and sensory exploitation. *Oxford Surveys in Evolutionary Biology*, 7:157–195, 1990.

[78] M. J. Ryan, J. H. Fox, W. Wilczynski, and A. S. Rand. Sexual selection for sensory exploitation in the frog *physalaemus pustulosus*. *Nature*, 343:66–67, 1990.

[79] N. Sanderson. Can gene flow prevent reinforcement? *Evolution*, 43(6):1223–1235, 1989.

[80] S. Schindler. Schemata and recombinations in genetic algorithms in particular with the existence of symmetries. Master's thesis, University of Leipzig, 2004. Original title in german: Schemen und Rekombinationsverfahren in Genetischen Algrotithmen insbesondere bei Vorliegen von Symmetrien.

[81] M. R. Servedio. Reinforcement and the genetics of nonrandom mating. *Evolution*, 54(1):21–29, 2000.

[82] M. R. Servedio and M. Kirkpatrick. The effects of gene flow on reinforcement. *Evolution*, 51(6):1764–1772, 1997.

[83] J. M. Setchell and A. F. Dixson. Changes in the secondary sexual adornments of male mandrills (*mandrillus sphinx*) are associated with gain and loss of alpha status. *Hormones and Behavior*, 39(3):177–184, 2001.

[84] T. B. Smith. Disruptive selection and the genetic basis of bill size polymorphism in the african finch *pyrenestes*. *Nature*, 363:618–620, 1993.

[85] E. Sober. *Philosophy of Biology*. Dimensions of Philosophy Series. Westview Press, 2000.

[86] M. Spence. Job market signaling. *The Quarterly Journal of Economics*, 87(3):355–374, 1973.

[87] H. T. Spieth. Courtship behavior in *drosophila*. *Annual Review Entomology*, 19:385–405, 1974.

[88] F. Spirito. The reduction of gene exchange due to a prezygotic isolating mechanism with monogenic inheritance. *Theoretical Population Biology*, 32(2):216–239, 1987.

[89] P. F. Stadler and C. Stephens. Landscapes and effective fitness. *Comments on Theoretical Biology*, 8(4–5):389–431, 2003.

[90] F. W. Stöcker and G. Dietrich, editors. *Concise Encyclopedia Biology*. deGruyter, 1996.

[91] K. Summers, R. Symula, M. Clough, and T. Cronin. Visual mate choice in poison frogs. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 266(1434):2141–2145, 1999.

[92] D. Udovic. Frequency-dependent selection, disruptive selection, and the evolution of reproductive isolation. *The American Naturalist*, 116(5):621–641, 1980.

[93] S. S. Utami, B. Goossens, J. R. Brutford, Michael W. de Ruiter, and J. A. van Hooff. Male bimaturism and reproductive success in Sumatran orang-utans. *Behavioral Ecology*, 13(5):643–652, 2002.

[94] G. S. van Doorn and F. J. Weissing. Ecological versus sexual seletion models of sympatric speciation: A synthesis. *Selection*, 2(1–2):17–40, 2001.

[95] J. G. van Rhijn. Behavioural dimorphism in male ruffs, *philomachus pugnax (l.)*. *Behaviour*, 47(3-4):153–227, 1973.

[96] S. G. Vandenberg. Assortative mating, or who marries whom? *Behaviour Genetics*, 2(2/3):127–157, 1972.

[97] B. A. Walther and D. H. Clayton. Elaborate ornaments are costly to maintain: evidence for high maintenance handicaps. *Behavioral Ecology*, 16(1):89–95, 2005.

[98] N. M. Waser and M. V. Price. Pollinator choice and stabilizing selection for flower color in *delphinium nelsonii*. *Evolution*, 35(2):376–390, 1981.

[99] C. Wedekind, T. Seebeck, F. Bettens, and A. J. Paepke. MHC-dependent mate preferences in humans. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 260(1359):245–249, 1995.

[100] P. M. West and C. Packer. Sexual selection, temperature, and the lion's mane. *Science*, 297(5585):1339–1343, 2002.

[101] L. L. Wolf. "Prostitution" behavior in a tropical hummingbird. *The Condor*, 77:140–144, 1975.

[102] T. K. Wood and M. C. Keese. Host-plant-induced assortative mating in *enchenopa* treehoppers. *Evolution*, 44(3):619–628, 1990.

[103] A. Zahavi. Mate selection - A selection for a handicap. *Journal of Theoretical Biology*, 53:205–214, 1975.

**Selbstständigkeitserklärung**

Hiermit erkläre ich, die vorliegende Dissertation selbständig und ohne unzulässige fremde Hilfe angefertigt zu haben. Ich habe keine anderen als die angeführten Quellen und Hilfsmittel benutzt und sämtliche Textstellen, die wörtlich oder sinngemäß aus veröffentlichen oder unveröffentlichten Schriften entnommen wurden, und alle Angaben, die auf mündlichen Auskünften beruhen, als solche kenntlich gemacht. Ebenfalls sind alle von anderen Personen bereitgestellten Materialien oder erbrachten Dienstleistungen als solche gekennzeichnet.

Leipzig, 25. Juni 2010
Susanne Schindler